Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

In vitro processing and complex formation of coronavirus polyprotein NSP7-10 region

DISSERTATION

with the aim of achieving a doctoral degree at the Faculty of Mathematics, Informatics and Natural Sciences

Department of Biology at Universität Hamburg

Boris Krichel
born in Wertheim am Main

Hamburg 2019

Reviewers

Dr. Charlotte Uetrecht

Heinrich Pette Institute,
Leibniz Institute for Experimental Virology, Hamburg, Germany

European XFEL GmbH, Schenefeld, Germany

Prof. Dr. Julia Kehr

Institute of Plant Science and Microbiology
Faculty of Mathematics, Informatics and Natural Sciences
Department Biology, Universität Hamburg, Germany

Submission of Thesis: May 29, 2019
Disputation: July 26, 2019
This version was printed for publication on August 22, 2019 with smaller editorial changes.

Heinrich Pette Institute
Leibniz Institute for Experimental Virology

This thesis was prepared from April 2014 to May 2019 under the supervision of Dr. Charlotte Uetrecht at the Heinrich Pette Institute, Leibniz Institute for Experimental Virology in the working group 'Dynamics of Viral Structures'. The second supervisor was Prof. Dr. Thomas Dobner, associated with the Department of Biology at the University of Hamburg and head of the research department 'Viral Transformation' at the Heinrich Pette Institute.

Gewidmet meiner Mutter und meinem Vater

Danksagung

Obwohl ich diese berufliche Laufbahn nie anvisiert hatte, wurde ich damals irgendwie gefunden, eingesammelt und mitgenommen. Deshalb, zuallererst, vielen Dank Charlotte für die letzten 5 Jahre. Du gabst mir die Chance viel zu lernen, zu erleben und die Massenspektrometrie zu lieben, was ich seitdem tue. Ich will mir bei Leibe nicht vorstellen, dass ich auch irgendwo anders hätte landen können.

Weiterhin möchte ich Prof. Thomas Dobner danken, der als Co-Supervisor für mich als Jungwissenschaftler zuständig war, der den Vorsitz der Prüfungskommission übernommen hat und dessen Türen im Labor, in der Kaffeeküche und im Büro immer für mich offenstanden. Ich danke auch Prof. Julia Kehr, Gutachterin dieser Thesis, und Prof. Hartmut Schlüter, die beide als Mitglieder der Prüfungskommission fungierten.

Insbesondere danke ich allen, die zum Abschluss dieser Arbeit beigetragen haben. Hervorheben möchte ich hier Prof. John Ziebuhr, Ramakanth Madhugiri und Ganesh Bylapudi für das Vertrauen und die wunderbaren Proben. Weiterhin möchte ich nennen Prof. Renato Zenobi und Martin Köhler für die Zeit an der ETH Zürich; Jan, Jerre und Nico von MS Vision; Jasmin, Marc, Simon und Daniel für Hilfe im Labor; Andi, Frederike, Charlotte, Ronja und Knut für Datenauswertung und Diskussion; Timothy Soh für Englischkorrektur. Das Coronavirus NSP7-10 Project wurde ursprünglich von Sven Falke und Prof. Lars Redecke begonnen, die mir besonders zu Beginn meiner Thesis mit Material und in Diskussionen eine Hilfe waren.

Während meiner Zeit als Doktorand durfte ich phantastische Studenten betreuen: Jonathan Zöller, Kira Schamoni, Cécile Demarez, Kinga Wrona, Daniel Woike und Simon Meyer. Danke! Jetzt noch einmal Grüße an alle Leute ohne die die Zeit am HPI nicht so gewesen wäre wie sie war. Grüße an alle vom Dobner Lab, Herker Lab, die ganze IT, die ganze Technik und die ganze Verwaltung. Unsere Juniorarbeitsgruppe Uetrecht (AG77, ftw!) soll auch an dieser Stelle gegrüßt werden: „Hallo Piggies, Kleinis, Cousins und andere Cavepeople. Ihr seid die Besten!“ Besondere Grüße gehen raus an meinen langjährigen Cavenachbarn und gleichgroßen Kuseng Dr. Johannes Heidemann, mit dem ich die beste Zeit hatte, und an Dr. Sarah Müncheberg, meiner kongenialen Partnerin zu unserer Zeit als Doktorandensprecher.

An Jana, Larissa, Mama, Michaela, Oma, Opa, Papa, Susi (alphabetisch ☺): Ihr gebt mir die Sicherheit, um zu sein. Danke! Weiterhin viel Liebe an alle in den Krichel und Geier Clans.

Last but not least! Die ganze Logistik zwischen der Dissertation und Familie war schon Wahnsinn! Das ging nur mit Hilfe meiner wunderbaren Frau Chrystal, die mich zur Promotion ermutigte, mich dabei unterstützte, und uns nebenbei zwei Töchter schenkte. Deshalb zum Schluss „Ich liebe euch meine schönen Girls Chrystal Lee, Aila Lee und Antonia Lee!“

Zusammenfassung

Coronaviren (CoV) infizieren unterschiedliche Spezies und verursachen schwere Anthropozoonosen im Menschen. Die CoV Polyproteine 1a/ab müssen prozessiert werden, um 11/16 Nichtstruktur-proteine (NSPs) freizugeben, die am Replikations-Transkriptions Komplex (RTC) beteiligt sind. Zwei essentielle Schritte dorthin sind (1) die Prozessierung der konservierten Region NSP7-10 von der Protease M_{pro} und (2) die daraufhin ausgelöste Komplexbildung der freigegebenen NSPs. In dieser Arbeit analysierte ich Prozessierung und Komplexbildung der CoV NSP7-10 Region mit Methoden der strukturellen Massenspektrometrie (MS).

In einer ersten Reihe von Experimenten untersuchte ich die Prozessierung von SARS-CoV NSP7-10. Zu Beginn wurden M_{pro} Spalteffizienzen mit FRET Peptid Substraten, mit analogen Spaltstellen zu SARS NSP7-10, bestimmt. Dann wurden relative Spalteffizienzen offengelegt, in einem Verfahren, das *in vitro* Prozessierung von SARS, 229e und FIP-CoV Volllängensubstraten mit nativer Massenspektrometrie kombiniert. Die Ergebnisse zeigen, dass es zwischen den Volllängensubstraten nur begrenzte Konservierung von Spalteffizienz gibt und wiedersprechen daher Peptid Analysen, die hier von mir und bereits vorher von anderen Wissenschaftlern durchgeführt wurden.

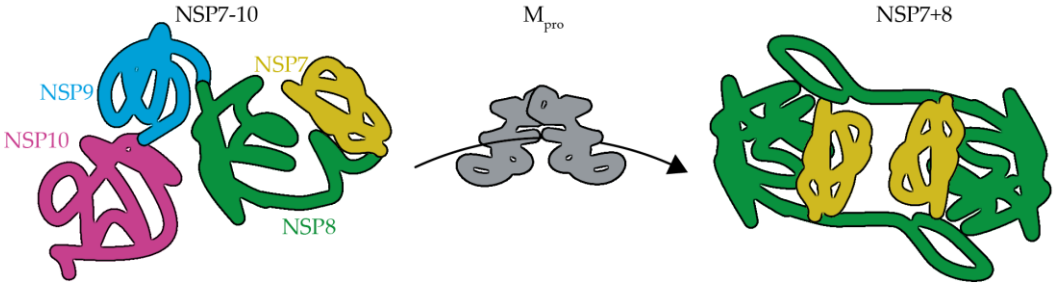
Mit einer zweiten Reihe von Experimenten erbrachte ich neue Erkenntnisse über bereits bekannte und noch unbekannte NSP7+8 Komplexe. Anfangs wurde das Muster ihrer nicht-kovalenten Interaktionen mit nativer MS abgebildet. Darin waren die prädominierenden Komplexe FIP und TGEV-CoV Hetero-Trimer NSP7+8 (2:1) und SARS, PEDV und 229e-CoV Hetero-Tetramer NSP7+8(2:2). Daraufhin wurden deren Gasphasendissoziationen mit kollisionsinduzierter Dissoziation mit Q-ToF Tandem MS getestet. Die Ergebnisse deuten darauf hin, dass die Tetramere von SARS eine andere Anordnung von Untereinheiten haben als die Tetramere von PEDV und 229e. Weiterhin dissoziierten die Komplexe von FIP und TGEV bemerkenswert ähnlich. Diese Resultate erlaubten die Validierung eines prognostizierten strukturellen Modells des TGEV Trimers.

In einer dritten Reihe von Experimenten benutzte ich Matrix-assistierte Laser-Desorption-Ionisierung (MALDI)-MS, um diese Prozesse aus einem anderen Blickwinkel zu betrachten. Die *in vitro* Prozessierung von hCoV-229e NSP7-10 wurde auch hier untersucht. MALDI-MS wurde auch mit chemisch vernetzten FIP und 229e-CoV Komplexen durchgeführt. Die gefundenen Prozessierung- und Interaktionsmuster stimmten mit den Ergebnissen der nativen MS überein.

In einem vierten und letzten Experiment untersuchte ich gespaltenes und ungespaltenes TGEV-CoV NSP7-8 mit Wasserstoff-Deuterium Austausch (HDX)-MS. Klare Unterschiede in der

Deuterium Aufnahme in spezifischen Regionen wiesen auf enorme Umstrukturierungen durch die Prozessierung hin. Die Visualisierung dieser Ergebnisse auf strukturellen Proteinmodellen förderte das Verständnis der Komplex-Bindestellen und erklärte die Instabilität der NSP8 Untereinheit, wenn sie vom Komplex isoliert vorliegt.

Zusammengefasst, unterschiedliche molekulare Prozesse der Prozessierung von CoV-NSP7-10 wurden mit Methoden der strukturellen Massenspektrometrie untersucht. Die CoV Spezies unterscheiden sich durch Spalteffizienzen in NSP7-10 und Komplexarchitektur von NSP7+8. Diese Ergebnisse erlauben neue Einblicke in Protein-Protein Interaktionen, welche als essentiell für Coronaviren gelten und könnten daher helfen, um antivirale Wirkstoffe zu entwickeln.



Volllängensubstrate
-Expression und Reinigung
-native MS

Prozessierung
-native MS
-MALDI-MS
-HDX-MS

Komplexbildung
-native MS
-CID
-XL-MALDI MS
-HDX-MS

Abstract

Coronaviruses (CoV) infect a wide range of species and can cause severe zoonotic disease in humans. The CoV polyproteins 1a and 1ab require processing to release non-structural proteins (NSPs) 1-11 and NSP 1-16 and form a replication transcription complex (RTC). It has been shown that two essential steps are (1) processing of the conserved polyprotein NSP7-10 region by main protease M_{pro} and (2) subsequent complex formation of the released NSPs. In this thesis, I analyzed processing and complex formation of the coronavirus NSP7-10 region with structural mass spectrometry.

In a first set of experiments, I studied processing of CoV NSP7-10. In the beginning, specific M_{pro} cleavage efficiencies were determined from FRET peptide substrates representing the SARS-CoV cleavage sites NSP7-10. Then, relative cleavage efficiencies were exposed, combining *in vitro* processing of SARS, 229e and FIP full-length protein substrates. The results show that within full-length substrates there is only limited conservation of cleavage efficiency and thereby oppose findings from peptide assays, of me here, and other researchers before.

In a second set of experiments, I provided novel insight into previously known and unknown NSP7+8 complexes of different CoV species. Initially, their landscape of non-covalent interactions was pictured by native MS. The predominant complexes found were FIP and TGEV-CoV hetero-trimeric NSP7+8 (2:1) and SARS, PEDV and 229e-CoV hetero-tetrameric NSP7+8(2:2). Thereon, gas-phase dissociation was tested with CID in Q-TOF tandem MS. The results suggest that tetramers of SARS have a different subunit arrangement than PEDV and 229e-CoV. Furthermore, the dissociations of FIP and TGEV-CoV were strikingly similar. From these results, a predicted model of the TGEV-CoV trimer was confirmed.

In a third set of experiments, I used high-mass MALDI MS for a different view on these processes. By revisiting *in vitro* processing of 229e NSP7-10, some open questions could be clarified. Carrying out MALDI-MS upon cross-linking of FIP and 229e-CoV NSP7+8, a similar complex landscape was observed as in native MS. In a fourth and last experiment, I compared uncleaved and cleaved TGEV-CoV NSP7-8 in HDX-MS. Clear differences in deuterium uptake in specific regions, indicated for tremendous structural rearrangement upon processing. Visualizing the results on structural models enhanced our understanding of binding sites in the complex and explained features of the polyprotein.

Summing up, multiple molecular processes upon processing of NSP7-10 were studied with structural MS. The CoV species vary in NSP7-10 cleavage sites efficiency and NSP7+8 complex architecture. These results provided new insights in protein interactions, which previously had been shown essential, and therefore could be helpful to develop antivirals.

Table of Contents

1 Introduction

1.1 Coronavirus

In 1965, the human coronavirus (hCoV) 229e was first isolated and described as a causative agent of upper respiratory tract infections. In fact, hCoV 229e and other hCoV species cause up to 30% of all cases of the common cold [1-5]. Due to this relatively harmless disease, coronaviruses had remained under the radar of many infection biologists.

Worldwide awareness of coronavirus as a pathogen came with the outbreak of severe-acute respiratory syndrome (SARS) in China in 2002/2003, initially transmitted from meat of palm civet cats to humans [6, 7]. The disease quickly spread to 29 countries accompanied by a high case fatality rate (~10%) [8]. At that time, scientists were unable to predict the course of the epidemic, because an understudied virus caused this unprecedented disease. Meanwhile, spectacular media coverage caused panic and major economic damage [9]. Eventually, SARS faded out through infection control and poor human-to-human transmissibility. No active disease has been detected since 2005, but in the end, SARS left 774 dead [8]. As a result, coronavirus research was suddenly of the highest interest and therefore largely funded [10].

In 2012, another wave of coronavirus awareness, research, and funding [11, 12] started with coronavirus infections causing Middle-East respiratory syndrome (MERS), which was initially transmitted from dromedary camels in Saudi Arabia [13, 14]. The disease spread quickly to other countries and even a new hotspot appeared in Korea [15]. This time, researchers and media had been better prepared. However, MERS has never faded out completely. Countries of the Arabian Peninsula report a couple of infections per month with an extremely high case fatality rate (~34.6 %) [16].

Both, SARS and MERS, are zoonotic infections, caused by cross-species transmission [17]. Their main reservoir are bats, but coronaviruses are further present in a wide range of wildlife, pets and livestock [18, 19]. In animals, coronaviruses are also known as major pathogens causing mostly gastrointestinal and respiratory, but also systemic disease. Some of these viruses are well-studied due to their economical relevance, such as porcine epidemic diarrhea virus (PEDV) and transmissible gastroenteritis virus (TGEV) resulting in a significant morbidity and mortality in piglets [20, 21]. Further interest was spurred by a virus formidably for pet lovers, feline intestine peritonitis (FIP); this disease, which once clinically manifested, results in the death of the cat [22, 23].

The versatility of coronavirus infection constitutes a risk that these viruses jump from animals over to humans and cause a severe epidemic, as it was the case for SARS and MERS. Currently, no therapy exists. More research is required to better understand pathology and molecular

biology of coronaviruses in order to help develop antiviral remedies, so that in case of a new outbreak, the world is maximally prepared.

1.1.1 Molecular biology of infection

The virion of coronaviruses has an envelope that is populated with a large protruding spike protein (S), which is responsible for the crown-like appearance and therefore the naming of the virus; in Latin, crown means corona (Figure 1). The particle further contains the membrane protein (M), small membrane protein (E) and nucleoprotein (N). N protects the positive sense single stranded RNA genome. Additionally, some CoV lineages possess a hemagglutinin-esterase (HE).

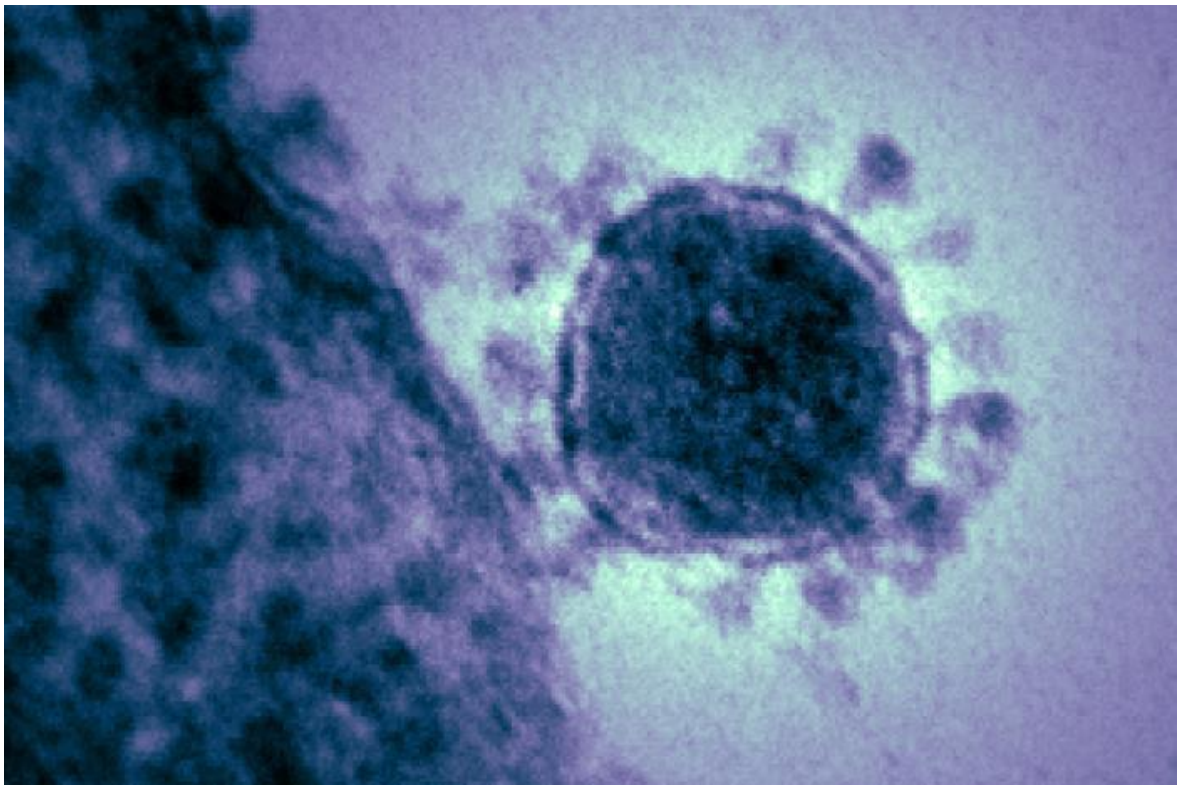


Figure 1**: Electron micrograph of MERS-virion.** Spike proteins protruding visibly from the viral particle and give it a crown-like appearance. Highly magnified, digitally colorized transmission electron microscopic (TEM) image reveals ultrastructural details exhibited by a single, spherical-shaped, Middle East respiratory syndrome coronavirus (MERS-CoV) virion. Free of license. creativecommons.org/licenses/by/2.0/. National Institutes of Health (NIH) Photo database National Institute of Allergy and Infectious Diseases (NIAID), 2014.

Most importantly is the spike protein, because the zoonotic potential of a coronavirus is directly dependent on the receptor binding properties of its protruding domain. The spike protein of SARS-like bat coronavirus is able to use receptors from humans, civets and Chinese horseshoe bats for cell entry [24] indicating the jump between species for SARS. Recently, molecular structures of S and its glycan shell were extensively studied by cryo-electron microscopy (cryo-EM) and mass spectrometry, thereby pushing the limits of understanding of host attachment

and membrane fusion [25-27].The infectious cycle begins with binding of the spike glycoprotein to its receptor where cleavage by host proteases triggers fusion activation with either the plasma membrane or endosomes [28]. Upon uncoating in the cytoplasm, a large part of the viral genome, open reading frames ORF1a/ab, is translated directly into replicase polyproteins 1a/1ab. These polyproteins require processing to release non-structural proteins (NSPs), which eventually form sites of replication by inducing a reticulovesicular network from the endoplasmic reticulum and therein assemble into a replication-transcription complex (RTC). In these sites of replication, mRNAs and structural proteins are produced while new virus particles bud into ER-Golgi intermediate compartments [29]. Finally, the virus containing vesicles fuse with the plasma membrane for release [30, 31].

1.1.2 Taxonomy

The international committee on Taxonomy of Viruses recognized and classified coronaviruses [32]. They belong to the order of *Nidovirales* as the family of *Coronaviridae*, which is closely related to *Arteriviridae*, another family of viruses known as pathogens of animals. According to different ORFs, *Coronaviridae* contains subfamilies *Alpha-*, *Beta-*, *Gamma- and Deltacoronavirus* (Figure 2). As human pathogens, *Betacoronaviruses* are well researched and within this subfamily, four separate lineages can be discerned, designated A through D. Differences between these lineages are for example the existence of an additional PL_{pro} domain in lineage A or a Hemagglutinin-esterase (HE) in lineage D.

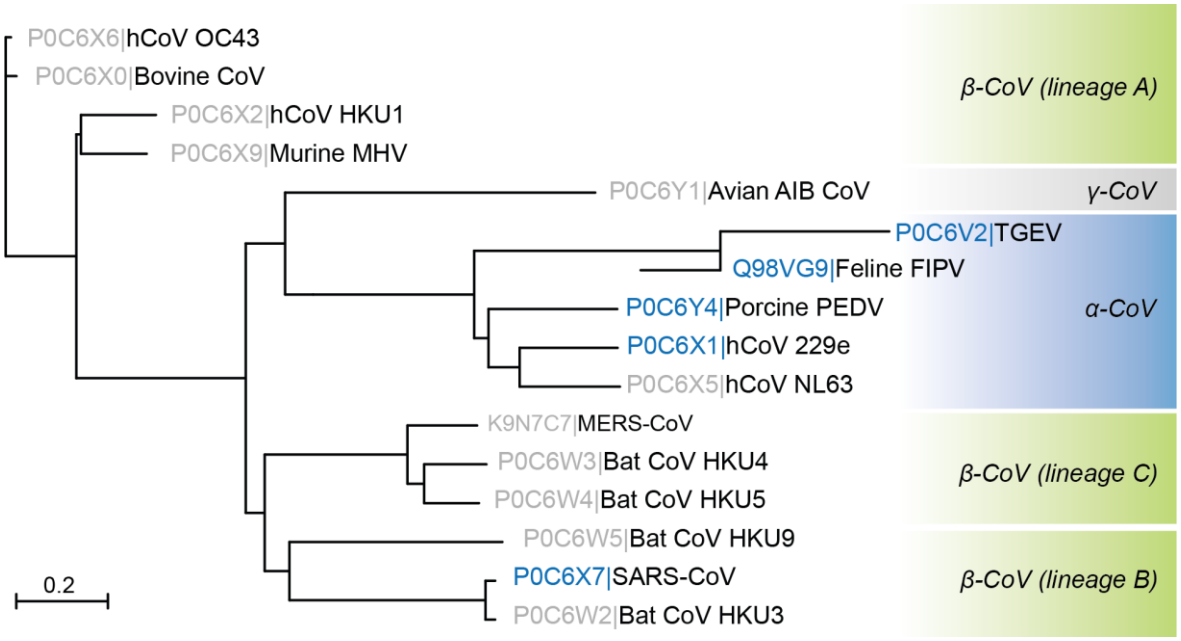


Figure 2: Phylogenetic tree of selected members of the subfamily *Coronavirinae*. The tree bases on amino acid sequences of replicase polyprotein 1ab, which is a topic of research in this work. Protein IDs indicated for proteins mentioned in this thesis (grey) or analyzed in this work (blue). Genera and lineages given (right). Phylogenetic tree generated with amino acid sequences retrieved from UNIPROT and used as input for the Virus Pathogen Resource online tool. Input parameters were the "Quick Tree" option using MUSCLE alignment and FastME approach [33].

1.1.3 Genome, replication and transcription

Coronaviruses have a ~32 kB genome, one of the largest known genomes of RNA viruses, and therefore, require special measures for effective translation and transcription. Only one third of the genome is reserved for all structural and accessory proteins. This region is different in between CoV species regarding length and number of genes. However, N, S and M genes are present in all known species. Two thirds of the genome contains a replicase gene ORF1a/ab, which is largely conserved between CoV species and encodes for the proteins for replication and transcription.

Initially, the replicase gene is directly translated from the positive-sense (+) single stranded RNA into either polyprotein pp1a (NSP1-11) or pp1ab (NSP1-16), depending on a ribosomal (-1)-frameshift [34] (Figure 3). Subsequently, the polyprotein undergoes processing by two internal proteases, the papain-like protease (PL_{pro}; NSP3) for NSP1-4 and the main chymotrypsin-like protease (M_{pro} or 3CL; NSP5) [35, 36] for NSP5-16. Processing is crucial, because it leads to maturation of the individual non-structural proteins (NSPs). Eventually, they take part in locating and forming the RTC, a membrane anchored but dynamic protein-RNA complex responsible for the replicative processes [37-40].

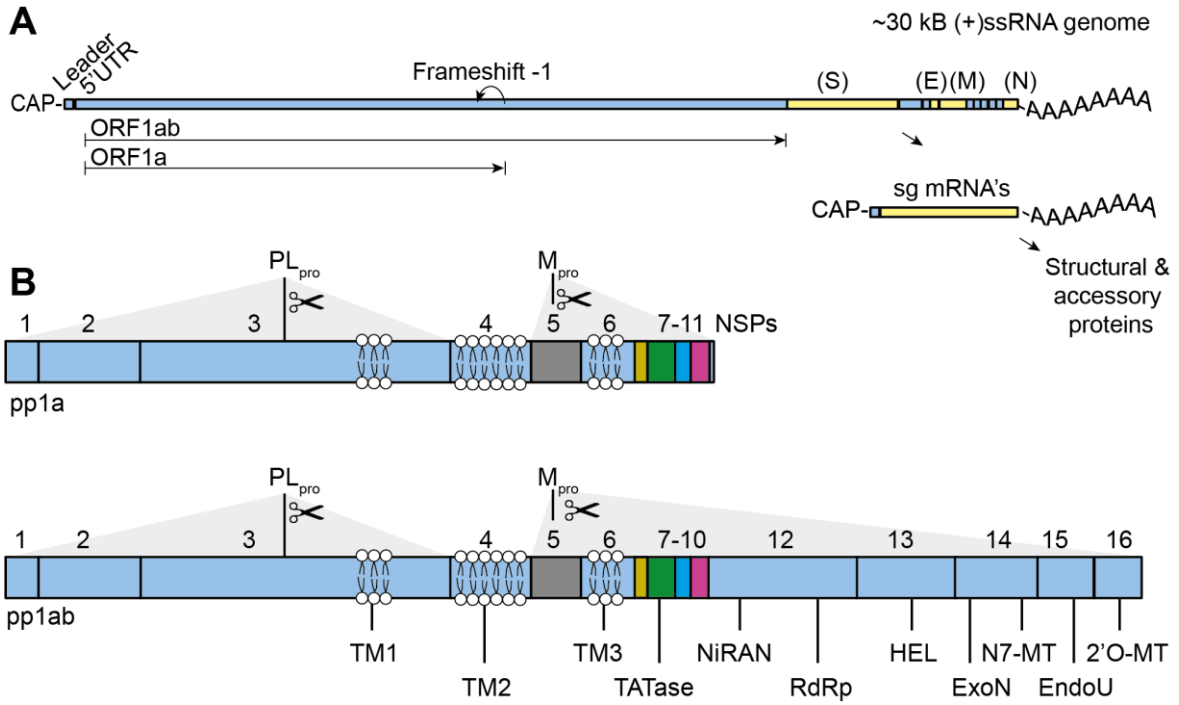


Figure 3: Genome and replication of coronaviruses. Schematic illustration shows SARS-CoV replication, which does not considerably differ from other CoV species. (A) CoV genomic RNA contains at the 5′ end two open reading frames ORF1a/ab, which are directly translated into pp1a/ab. Indicated in yellow are genes for structural proteins, present in all species of *Coronaviridae*. Sequences at the 3′ end act in *cis* with the 5′ UTR to form ORFs and facilitate transcription of sg mRNAs. (B) Replicase pp1a/ab with NSP1-11/1-16. The NSP7-10 region is highlighted (colored) and a topic of research in this work. TM1, TM2 and TM3 are transmembrane regions. Enzymatic activities indicated are valid for pp1a and pp1ab. Two protease and their cleavage sites indicated (scissor symbols, grey areas).

The RTC locates in proximity to translation initiation factors at virus induced double membrane vesicles (DMVs) that serve as protective workspace and niche for the replicative processes [41-44]. Transmembrane domains are present in NSP3, NSP4 and NSP6, inducing a larger reticulovesicular network from intracellular membranes such as the endoplasmic reticulum (ER). However, NSP3 and NSP4 are required and sufficient for the formation of DMVs.The RTC facilitates versatile functions, most importantly, replication of genomic RNA via synthesis of the negative (-) strand as well as transcription of a set of subgenomic mRNAs (sgRNA) via the

cis-acting RNA leader sequence at the 5′ end of the genomic RNA (Figure 4). For translation and host evasion, the RTC further equips sgRNAs with a 5′cap and a 3′ polyadenylation. The sgRNAs encode for structural and accessory proteins and vary between coronavirus species [45-47].

One key activity of the coronaviral RTC resides in the RNA dependent RNA polymerase (RdRp, NSP12) and the exoribonuclease (ExoN, NSP14), a proofreading enzyme with additional function as a (guanine-N7) methyl transferase (N7-MTase) for mRNA capping. The interplay between NSP12 and NSP14 is crucial for high-fidelity replication of the exceptional long genome.

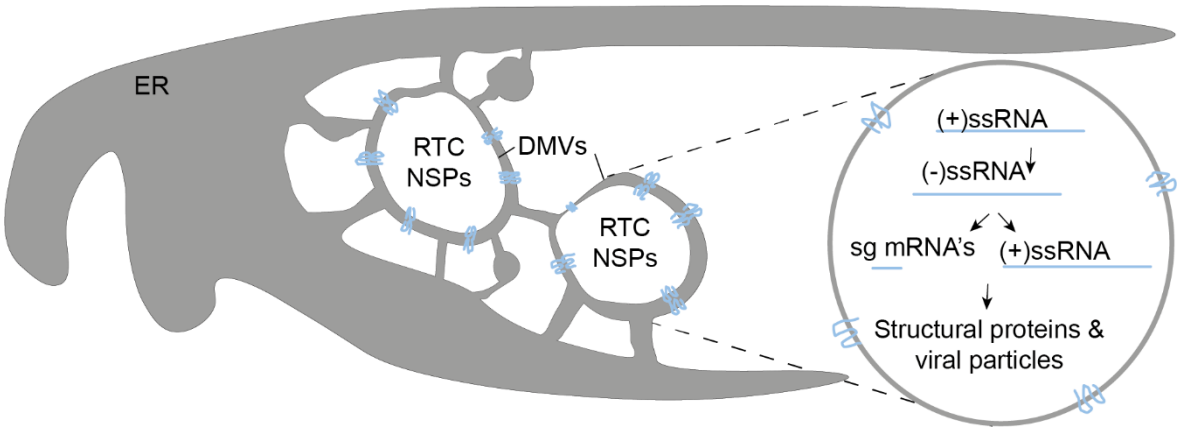


Figure 4: Location of replication-transcription complex. Nsp induced reticulovesicular network in endoplasmic reticulum (ER). Double membrane vesicles (DMVs) used as site of replication, containing the nsp domains forming the replication transcription complex (RTC). The RTC facilitates the synthesis of the negative strand, transcription of sg mRNAs and (+)ssRNA, which are required produce structural proteins and viral particles.

The so-called core enzymes reside in NSP12-NSP16, encompassing further RNA capping a helicase (NSP13), endoribonuclease (NendoU, NSP15) and 2′-O–methyl transferase (2′O-MTase, NSP16). The current unsolved matters regarding the RTC are the role of host proteins and the other NSP proteins (e.g. NSP2, NSP7-10), which locate in proximity of the DMVs, as well as the identification of an RTC membrane anchor [48]. For these and other reasons, the structural assembly of the RTC remains elusive [45-47].

Upon processing of the NSP7-10 region by Mpro the four small regulatory proteins NSP7, NSP8, NSP9 and NSP10 are released and form complexes with core enzymes. Some remaining open questions of these processes are addressed in this thesis and therefore, to provide understanding, current knowledge is reviewed in the next paragraphs.

1.1.4 Main protease M_{pro}

M_{pro} is responsible for polyprotein processing at NSP inter-domain junctions between NSP4-16. This protease is indispensable for viral progeny and homologs are already targets for antiviral therapy. Therefore, the structure and function of M_{pro} has been extensively investigated [36, 49-51] and reviewed [52].

In early infection or right after translation, M_{pro} is part of the viral polyprotein as immature NSP5, located between the two membrane domains NSP4 and NSP6. To mature from the polyprotein, the protease is released in two main steps. First, the N-terminal junction is *cis*-cleaved by a transient dimer of immature NSP5 [53, 54]. Second, the C-terminal junction is *trans*-cleaved by another transient or matured dimer of NSP5, freed from the N-terminus [36]. Only NSP5 with authentic N- and C-termini can become a fully active M_{pro} dimer. There are indications that after infection it could take hours for M_{pro} to exhibit full protease activity and facilitate processing of the polyprotein [55].

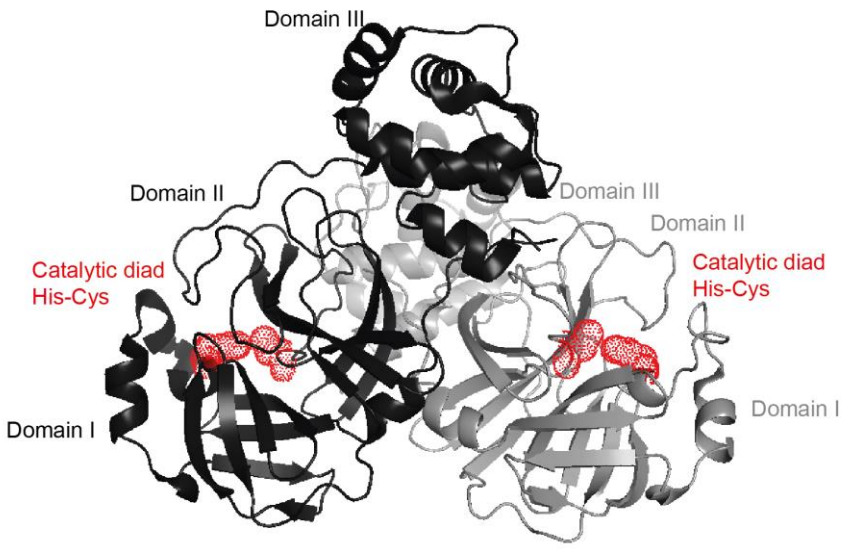


Figure 5: Molecular structure of authentic SARS Mpro homo-dimer. Shown are two protomer units (grey, black) with highlighted electron density of amino acids of the catalytic His-Cys dyad (red). The substrate-binding pocket locates in proximity to the catalytic dyad between domain I and domain II. Structure published[56] and available at PDB-ID: 1UK3. Modifications made with PyMol.

The structure of M_{pro} is a homo-dimer of about 70 kDa with three larger domains per protomer [51, 57] (Figure 5). In SARS-CoV M_{pro} [58], domain I (residues 8–101) and II (residues 102–184) have an antiparallel ß-barrel structure, which is similar to other CoV proteases and reminiscent of the chymotrypsin (3CL)-like serine proteases. Domain III (residues 201–303) contains five α -helices and is connected with domain II over a long loop region (residues 185–200). The active site consists of a catalytic dyad (Cys145 and His41), and the substrate-binding site is located in a cleft between domains I and II. In solution, the protein is in a monomer-dimer equilibrium.

However, particularly one structure by X-ray crystallography shows an octamer, which was proposed to not lose its activity at low concentrations, such as at the beginning of an infection during protease maturing [59].

The dimer dependent activity of M_{pro} is a largely investigated matter [36, 50, 52, 60-66]. In general, wild type enzymes are most active and dimerization mutants are inactive. The activity depends on the monomer-dimer equilibrium, and thus on concentration. The K_D of M_{pro} dimerization was often determined experimentally but reported values range from 0.35 nmol/L[66] to 227 μmol/L[65]. The protease's properties are influenced by N- and C terminal-fusion tags that inhibit dimerization and by presence of substrates that promote dimerization [67, 68].

Substrate specificities of M_{pro} were previously studied with peptide libraries and *in silico* approaches [69, 70]. Most efficiently cleaved by M_{pro} are ideal peptide sequences that resemble the auto-cleavage site NSP4-5 [71, 72] (Table 1).

Table 1: Preferred substrate specificity of M_{pro} in different CoV. Substrate specificities in *alphacoronavirus*, hCoV-NL63, *betacoronavirus*, hCoV-OC43 and SARS-CoV, as well as *gammacoronavirus*, Infectious bronchitis virus (IBV) at position P5-P3' found from mutations of the NSP4-5 auto-cleavage site sequence [71, 72]. Amino acids given in single letter code. Ideal sequence deviates at P4 with NL63, OC43 and SARS prefer alanine (A) while IBV prefers proline (P).

Pos.	SARS NSP4-5	ideal	most preferred residue
P5	S	V	high β-sheet propensity
P4	A	A/P	small hydrophobic res.
P3	V	R	pos. charged res./ high ß-sheet propensity
P2	L	L	hydrophobic res. without ß-branch
P1	Q	Q	Q
P1'	S	S	S/A/C/ small res.
P2'	G	G	S/no strong pref
P3'	F	F	R/no strong pref

1.1.5 M_{pro} FRET peptide assays

Since it is a potential drug target, many researchers and pharmaceutical companies tackled the inhibition of M_{pro} by peptides or small molecules targeting dimerization but mostly directly the active site. However, currently no active pharmaceutical ingredient (API) has been commercially launched in a larger scale.

One common method in screening for M_{pro} activity-inhibiting molecules is a protease activity assay with aforementioned ideal peptides substrates labelled with a Förster resonance energy transfer (FRET) fluorophore pair. These peptides are labelled at N- and C-terminus with two fluorophores, having overlapping emission to absorption wavelength, and therefore function as fluorescence donor to acceptor pair. Upon excitation, the donor absorbs light and transfers

energy through a non-radiative mechanism. Thereby, fluorescence at donor emission wavelength is quenched.

This principle relies on the proximity of donor and acceptor. Upon cleavage of the peptides, the donor and acceptor at C- and N-terminus are separated and changing fluorescence intensity can be recorded. This principle was adopted for high-throughput screening because in recent years, custom FRET peptides became reasonably priced as well as that new fluorophores have higher brightness and solubility.

Enzyme inhibiting screening assays like these, are often performed at concentrations lower than substrate saturation. Nevertheless, in screening for an inhibiting molecule, merely the apparent enzymatic parameters are important, which can be easily determined in comparative assays. For example, to determine the relative cleavage efficiency of a protease, first the efficiency without inhibitor is determined, which is considered as 100%, such as M_{pro} with an ideal FRET peptide substrate. Then, efficiency is tested in other conditions, such as such as the Mpro with an ideal FRET peptide substrate plus inhibitors, and the results are set in relation to the 100% value.

To determine relative efficiency between different FRET substrates a similar approach is performed. However, beforehand a little bit more effort is needed in determining additionally substrate specific values such as fluorescence coefficient. These and other peptide assays have already been performed to measure relative substrate efficiencies of M_{pro} at different coronavirus cleavage sites and thereby, processing was studied.

1.1.6 Polyprotein processing

The NSP7-10 region of the coronavirus polyproteins pp1a/ab contains four relatively small NSP domains of about 10-21 kDa that are important for full activity of the RTC. Processing of NSP7-10 is facilitated at cleavage sites, which reside at inter-domain junctions. The order of processing is directly dependent on the M_{pro} specific activity at these cleavage sites. Therefore, a deviating substrate specificity could be an important regulatory mechanism for timely coordinated release of NSP7, NSP8, NSP9 and NSP10 and thus full functionality of the RTC.

Within the polyprotein, the regulatory domains have a distinctive position in pp1a/ab. They locate in between the membrane domain NSP6 and either NSP11, a short peptide sequence of 6-13 amino acids NSP11, in pp1a or NSP12-16, the core enzymes, in pp1ab. Due to the less frequent frameshift, pp1a is present in excess and the C-terminus of NSP7-10 region is mostly connected to NSP11 (Figure 6).

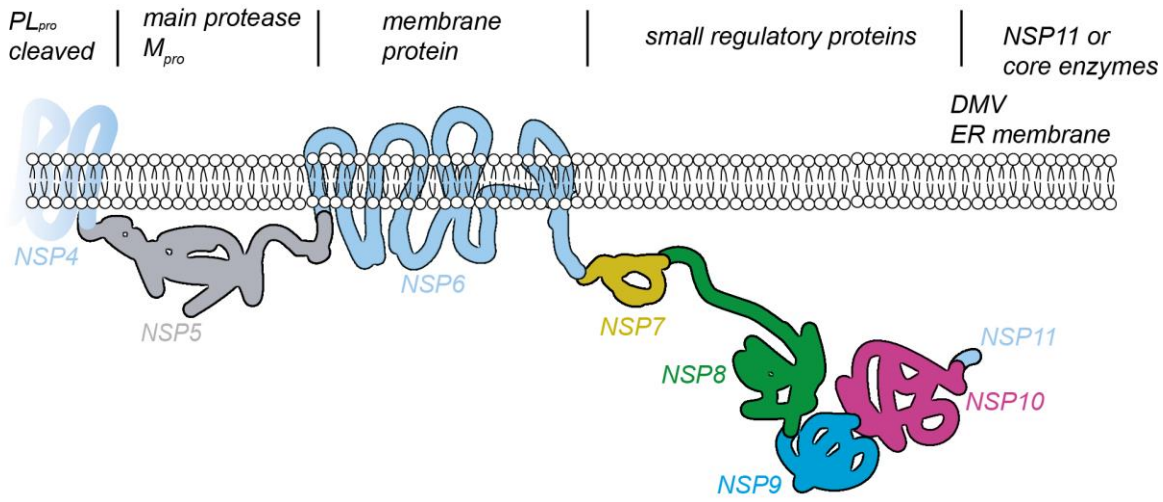


Figure 6: Location of regulatory region NSP7-10 within pp1a/ab. Depiction of one possible scenario as implicated by current research. At the beginning of infection, immature NSP5 cannot exhibit activity as the fully matured M_{pro}. Therefore, polyprotein processing is slow, and uncleaved polyprotein might be membrane anchored. There it could act as a *trans*-binding scaffold for core enzymes or as a functional precursor. Processing would constitute a regulatory switch, such as from negative to positive strand synthesis.

The possibility exists, that the NSP7-10 region is anchored in the ER by its N-terminus, the membrane protein NSP6, which is part of the reticulovesicular ER network. A molecular structure of NSP6 or interactions with other NSP domains remains elusive. However, *in silico* predictions suggest five transmembrane domains and a small soluble cytosolic endo-domain (6 kDa) at its C-terminus, where cleavage at the NSP6-7 site occurs [73]. The relative efficiency of M_{pro} at the NSP6-7 cleavage site was determined with artificial peptides. They found that

relative cleavage efficiency of this site is equal to other sites in MERS but extremely low in SARS [74, 75].

Schiller et al. performed pulse chase experiments in MHV infected cells and found that a 150 kDa protein encompassing NSP4-11 is present on a timescale of hours when other regions of the polyproteins are already cleaved [55]. These results indicated a long lasting and membrane anchored intermediate NSP4-10, including NSP7-10.

Some studies offer information about fate of the NSP7-10 region during infection in cell culture. Sawicki et al. performed a genetic and functional analysis of MHV mutants [76]. They showed that pp1a forms a single complementation group NSP1-11. They concluded that pp1a encodes a functional polyprotein or cis-acting protein complex that acts as a binding scaffold for the core enzymes. They also suggested that processing of pp1a could be an important switch for viral synthesis from minus to plus strand.

In another study, Tan et al. suggested that the actual protein NSP7-10 is a longer lasting intermediate product that interacts with NSP12 from infectious bronchitis virus (IBV) *in vitro* [77]. Even though this investigation did not directly target the processing, it showed that the NSP7-10 region exists during infection and that it can act as a binding partner for a core enzyme.

In one particularly revealing study, Deming et al. gave insight into details of NSP7-10 processing with a MHV reverse genetics system [78]. The investigators reported that deletion of single NSP domains of NSP7-10 leads to unviable virus. Furthermore, rearrangement of NSP7-8 to NSP8-7 and mutation of the NSP7-8 or NSP8-9 cleavage sites gave an unviable virus. Other than mutation of the NSP9-10 cleavage site that gave viable. These results highlight that processing of NSP7-10 is critical for virus progeny and further suggest that at one point during replication, intermediate proteins could be needed, such as an unperturbed, correctly arranged NSP7-8 or an NSP9-10 protein.

Previously, the locations of cleavage sites of coronaviruses were predicted *in silico* and tested *in vitro* [47, 55, 79, 80]. Siddel and Ziebuhr first proved experimentally that cleavage sites and products in NSP7-10 exist using recombinant proteins of different length from hCoV 229e NSP4-12 [81]. Continuing, they revealed that full-length NSP7, NSP8, NSP9 and NSP10 are present in virus-infected cells.

Relative efficiencies were determined in several studies in M_{pro} assays using peptide substrates representing the cleavage sites (Table 2). Hegyi and Ziebuhr used artificial peptides (15 aa, P8-P7′) in competitive protease assays with an HPLC readout. They reported that there is substrate conservation between three coronaviruses, MHV, 229e and TGEV and a similar relative substrate specificity between cleavage sites [82]. Particularly, the NSP8-9 site stood out because

it had significantly reduced substrate efficiency but was highly conserved as a non-canonical cleavage site NNE(L/I)MP. The conservation lead them to speculate that either the slow cleavage of the NSP8-9 site is preserved during evolution, to extend the half-life of a precursor protein NSP8-9, or that the N-terminus of NSP9 (NNE-motif) is required for the biological activity.

Table 2: Relative substrate efficiencies of peptides representing polyprotein 1a cleavage sites. The relative substrate efficiencies were determined with peptides in different assays. [a] Data for 229e, TGEV and MHV determined with peptide (15 aa, P8-P7′) protease assay and a HPLC readout from Hegiy and Ziebuhr [82]. [b] Data from SARS determined with peptide (11aa, P6-P5′) protease assay and a HPLC readout from Fan et al. [74]. [c] Data from MERS determined with DABCYL and EDANS labelled peptide (20aa, P12-P8′) protease assay with fluorescence readout [75].

cleavage site	species	P5	P4	P3	P2	P1	P'1	P'2	P'3	relative substrate eff.[a,b,c]
NSP4-5	229e	G	S	T	L	Q	A	G	L	1.00
	TGEV	N	S	T	L	Q	S	G	L	1.00
	MHV	T	S	F	L	Q	A	G	K	1.00
	SARS	S	A	V	L	Q	S	G	F	1.00
	MERS	S	G	V	L	Q	S	G	L	1.00
NSP6-7	SARS	V	A	T	V	Q	S	K	M	0.02
	MERS	V	A	A	M	Q	S	K	L	0.28
NSP7-8	SARS	R	A	T	L	Q	A	I	E	0.05
	MERS	P	S	V	L	Q	A	T	E	1.6
NSP8-9	229e	V	V	K	L	Q	N	N	E	0.15
	TGEV	T	T	K	L	Q	N	N	E	<0.05
	MHV	T	V	V	L	Q	N	N	E	<0.05
	SARS	A	V	K	L	Q	N	N	E	0.02
	MERS	A	V	A	L	Q	N	N	E	0.30
NSP9-10	229e	T	V	R	L	Q	A	G	K	0.64
	TGEV	T	V	P	L	Q	A	G	K	0.39
	MHV	T	V	R	L	Q	A	G	T	0.26
	SARS	T	V	R	L	Q	A	G	N	0.22
	MERS	T	V	R	L	Q	A	G	S	0.71

Other investigators also relied on artificial peptides to determine M_{pro} activity at polyprotein cleavage sites [60]. Fan et al. analyzed M_{pro} activity for SARS-CoV cleavage sites from artificial peptides (11aa, P6-P5′). They found that relative substrate efficiencies $(v_{max}/k_m)_{rel}$ of SARS are similar to 229e, MHV and TGEV. By comparing secondary structure of these peptides with substrate activity, they revealed that β-sheet-like structures tend to react faster.

In a similar study [75], but using fluorescence as a readout, Wu et al. analyzed the substrate activity of MERS-CoV from artificial peptide substrates (20aa, P12-P8′) labelled for Förster-resonance energy transfer (FRET) with N-terminal DABCYL and C-terminal Glu-EDANS. They reported that the relative efficiency of auto-cleavage site NSP4-5 is lower than expected, which

could be due to a Pro in the P3 position hindering cleavage, as identified previously [71]. Even though their determined relative efficiencies were not as variable, results indicate that NSP8-9 is one of the poorer substrates poor substrate, which is in line with the other studies.

Discussing the artificial peptides, Siddel and Ziebuhr noted that besides the primary structure, the polyprotein conformation might also contribute to the accessibility of specific cleavage sites [81]. Therefore, full-lengths proteins could display a more native-like substrate that considers the three-dimensional conformation.

Other than the molecular structure of isolated NSP domains, there is no knowledge about the structure of CoV's polyprotein. Therefore, the influence of polyprotein structure on cleavage efficiency is difficult to estimate. Two scenarios are possible. Firstly, if the polyprotein is an array of NSPs arranged like 'pearls on a string' separated by a flexible linker, then the cleavage site and the cleavage efficiencies might be similar to peptides. Secondly, if an overall structure of the polyprotein exists that is different from the structure of the NSP proteins, then the cleavage efficiencies might be considerably different.

In his PhD thesis, Sven Falke performed structural analysis of full-length NSP7-10 and showed overall structure of NSP7-10 by small angle X-ray scattering [83]. He also performed *in vitro* processing of NSP7-10 with SDS-PAGE readout and the results suggest a timely coordinated processing reaction, which was not further analyzed. Currently, a more detailed study on polyprotein processing with full-length NSP substrates has not been published.

The structures and interactions of the matured NSPs are well studied; a summary thereof follows in the next sections.

1.1.7 Complex formation of the regulatory non-structural proteins

1.1.7.1 NSP7+8 – role in polymerase function

The NSP7 and NSP8 are two interacting proteins, which act as processivity factor of the RNA-dependent RNA-polymerase (RdRp) residing in NSP12 (See Figure 7). Molecular structures show NSP7 (~9.5 kDa) as a compact protein of α-helices and NSP8 (~22 kDa) in a golf-club-like shape with a C-terminal compact head domain and a N-terminal α-helical elongated shaft [84-86]. Previous reports consistently show binding of NSP7 to the region between the head and shaft domains of NSP8 via mainly hydrophobic interfaces and two or three hydrogen bonds. Nevertheless, there is no consistency in the reported stoichiometry of NSP7+8.

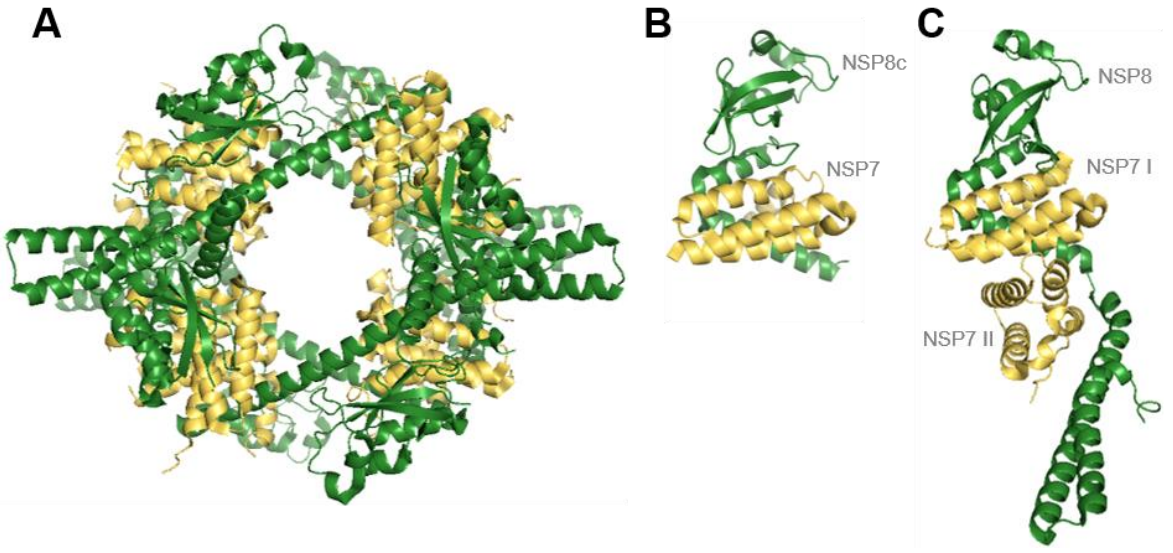


Figure 7: Previously solved molecular structures of NSP7+8 complex. Molecular structures solved with X-Ray crystallography of NSP7 (yellow) and NSP8 (green) in complex. **(A)** SARS NSP7+8 (8:8) hexa-decamer (PDB-ID: 2AHM [86]). **(B)** SARS NSP7+8 (1:1) hetero-dimer with a chopped NSP8 (aa1-78) (PDB-ID: 5F22) [87]) **(C)** Molecular structure of FIP NSP7+8 (2:1) hetero-trimer (PDB-ID: 3UB0 [84]) with different stoichiometry shows aa1-191.

Molecular structures of SARS NSP7+8 have been solved as a hexa-decamer and as a tetramer [86, 87]. Zhai et al. reported a hexa-decamer (8:8) that displays a hollow, cylinder like structure assembled from four tetrameric sub-complexes, termed T1 and T2, which structurally differ in folding of the elongated N-terminal α-helix of NSP8, which is either straight or bent with a loop. The tetramers interwove via the NSP8 elongated domains to form the hexa-decamer ring. A positive electric potential of the hollow channel and electric mobility shift assays imply RNA binding capacity. A detailed analysis of the NSP7+8 structure in solution has not yet been performed but Te Velthuis et al. observed a larger complex in a size exclusion chromatogram, possibly portraying the hexadecamer stoichiometry [88].

A ‘resectioned’ SARS-CoV NSP8 from cleavage of an unknown protease was reported by Li. et al. [87]. They roughly located the ‘resectioning site’ to the NSP8 N-terminal loop that has conserved amino acid residues in CoV. Continuing, they engineered a truncated C-terminal NSP8 (NSP8c) and solved the structure, together with NSP7, as a NSP7+8(1:1) dimer showing that binding of NSP8c to NSP7 is still possible. Finally, they speculated that ‘resectioning’ could also happen *in vivo* and serve as a mechanism for tuning viral replication.

A homologous FIP-CoV complex solved from X-ray crystallography and reported by Xiao et al. displays a different stoichiometry of NSP7+8 (2:1) [84]. The monomeric constituents have a similar structure, but while one NSP7 (NSP7I) is binding between the head and shaft domains of NSP8, as in the SARS complex, the other NSP7 (NSP7II) is binding to NSP7I and the NSP8 shaft. Altogether, sequences and binding interfaces from SARS and FIP NSP7+8 are conserved between them and the reason behind different stoichiometries remains elusive.

The function of NSP7+8 complexes is well studied. Several researchers reported a weak primer extension, suggesting a functional role of it as a second non-canonical polymerase [89], or *de novo* initiation, suggesting a function as a primase essential for NSP12 RdRp [84, 89, 90]. Newer reports questioned a clear division of primase and RdRp activity and rather favored NSP7+8 to serve as an essential processivity factor for the RdRp NSP12.

The essential interaction of NSP7+8+12 regarding polymerase functionality was described by Subissi et al. [91]. They showed that highest polymerase activity is reached upon incubation of NSP12 with either pre-mixed NSP7+8 or a covalently bound NSP7---8, connected via a flexible linker. From these results, it appears that complex formation of NSP7+8 is a prerequisite of NSP12 polymerase activity.

Recently, Tvarogová et al. reported that a recombinant form of hCoV229e, SARS and FIP NSP8, alone, has a metal ion-dependent RNA 3'-terminal adenylyl transferase (TATase) activity, which led them to speculate, that it could be involved in synthesis of the 3’polyA tail of the CoV (+)ssRNA [92].

Recently, a cryoEM study by Kirchdoerfer et al., revealed the structure of NSP7+8 in complex with the NSP12 polymerase (complex ~160 kDa) [93] (PDB file not yet published). The structure of NSP12 (~120 kDa) shows a polymerase domain resembling a cupped right hand similar to other polymerases and a nidovirus-unique N-terminal extension, termed NiRAN (nidovirus RdRp-associated nucleotidyl transferase) containing two Zinc atom binding sites. In complex with NSP7+8, the stoichiometry is unexpected resulting in NSP12:(NSP7+8):NSP8 of 1:1:1. The manifold interactions in this complex show the versatility of the NSP7+8 domains in combination with core enzymes, which are far from being understood.

NSP7+8+12 can also interact with the proofreading complex NSP10+14. For enzymatic activity of this complex, NSP10 is indispensable.

1.1.7.2 NSP10 - a co-factor for proofreading and capping

The structure of NSP10 (~15 kDa), solved as a monomer or dodecamer, exhibits a mixed α/β-fold and two Zinc finger domains, so far belonging to a structural family uniquely found in coronaviruses [94-96] (**Figure** 8). Interaction of NSP10 with one of the core enzymes either NSP14 or NSP16 is required for their activation and previously studied in detail [97, 98].

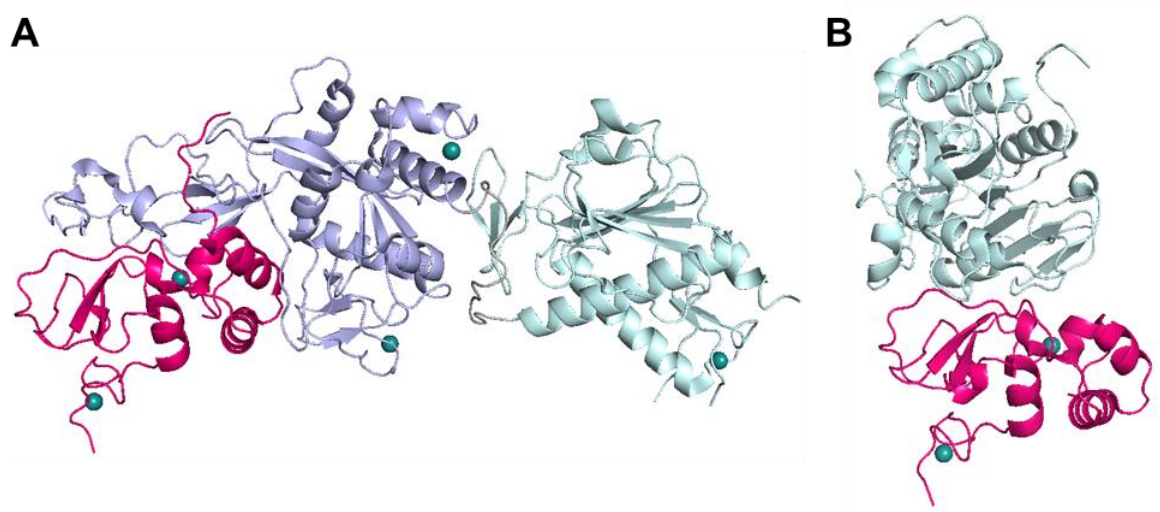


Figure 8: Complexes formed of NSP10 with coronavirus core enzymes. Molecular structures of NSP10 (pink) in complex with core enzymes. Zinc binding sites shown (deepteal). Structures modified with PyMol. **(A)** NSP10+14 is a bi-functional enzyme complex. ExoN domain (purple) and 2'O MTase (teal) are connected via a flexible hinge domain (grey) and have independent functions. NSP10 enhances ExoN, but appears not to influence structure of function of 2'O MTase. (PDB-ID: 5NFY [99]) **(B)** NSP10+16 is a N7 MTase absolutely requiring NSP10 for enzymatic activity and stability in vitro. (PDB-ID: 2XYR [98].

NSP14 (~60 kDa) is a bi-functional enzyme with N7-Methyltranferase (MTase) at its C-terminus and Exoribonuclease (ExoN) at its N-terminus [100]. Upon binding of NSP10, the ExoN activity of NSP14 is enhanced >35 fold, strictly targeting dsRNA excising in 3'-5' direction [97]. In fact, ExoN has proofreading function that is most likely required for replicating large genomes such in CoVs [101-103]. Most impressively, Ferron et al. reenacted the interplay between the polymerase complex and NSP10+14 *in vitro*. They observed that Ribavirin, an antiviral nucleoside analog, is first incorporated by the polymerase, but then excised by NSP10+14 [99] [91]. They further reported that both the ExoN and N7-MTase domains of NSP14 have an interface with the polymerase complex, but not NSP10.

Other than ExoN, the N7-MTase domain of NSP14 is not influenced, structurally or functionally, upon binding of NSP10. Nevertheless, in capping of the viral RNA, the function of N7-MTase directly precedes the function a 2'O MTase in NSP16, which absolutely requires binding of

NSP10 [98, 104, 105]. Decroly et al. reported a NSP10+16 (~38 kDa for NSP16) complex structure [106]. They suggest a possible binding site for interaction between both complexes, NSP10+14 and NSP10+16. Such a complex could be beneficial in spatially coordinating capping with methyl transfer.

1.1.7.3 NSP9 – a RNA binding protein

Compared to other proteins of the NSP7-10 region, no enzymatic function of NSP9 has been assigned. In MHV-A59, NSP9 localizes to late endosomes at sites of viral replication, together with NSP7, NSP8, and NSP10, suggesting that it is likely to play a role in the replication complex [107].

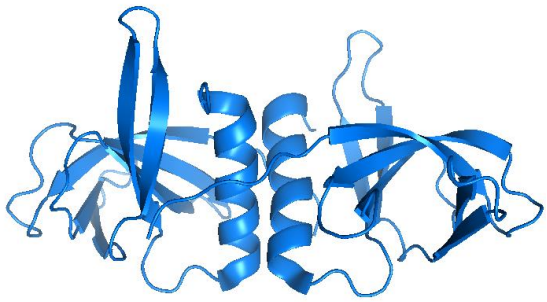


Figure 9: Helix-helix dimer of SARS NSP9. Molecular structure showing the parallel helix-helix dimer of SARS NSP9 connected via hydrophobic interactions. Structure modified with PyMol (PDB-ID: 1QZ8 [108].)

Multiple studies reported a monomer-dimer equilibrium and nucleic acid binding ability of NSP9. The protein-RNA binding affinities were determined with many different techniques and generated a large range, ~1-30 μM K_D [107, 109, 110] [111]. Mobility shifts are more dramatic with RNA than with single-stranded DNA under the same conditions [112]. Structures from X-Ray crystallography show that NSP9 has a single alpha helix fold and forms dimers in different modes [113, 114] (**Figure** 9). However, more recent studies indicate that proper dimerization is facilitated by hydrophobic interactions of two parallel alpha helices in a GXXXG motif [115]. A mutation of a single residue (G104E) within this motif terminates viral assembly *in vivo* [111]. The attributed properties suggest an essential RNA binding function of the NSP9 dimer in the coronavirus replicative organelles.

The inter-domain junction NSP8-9 has a non-canonical cleavage site that is conserved between CoV species. Particularly different is the highly conserved NNE motif [116] (P1'-P3') at the N-terminus of NSP9, while other NSP domains have aliphatic residues at this position. The purpose of the NNE motif has not yet been determined.

1.2 Structural mass spectrometry

Most of the experiments described in this thesis relied on methods of structural mass spectrometry. In order to understand them, a basic knowledge of mass spectrometry is beneficial. Therefore, this chapter properly introduces the techniques used. Partially included in this introduction are two mass spectrometry reviews, which I co-authored during my time as a PhD student. These are *Mass Spectrometry Goes Viral,* Dülfer, Kadek, Kopicki, Krichel and Uetrecht.; under submission (2019) (Sections 1.2.1, 1.2.2, 1.2.2.1, 1.2.5 and 1.2.8, Figure 12 and Figure 13) [117] and *Native Massenspektrometrie für die Proteinstrukturanalytik,* Heidemann, Krichel and Uetrecht; Biospektrum (2018.) (Figure 10 and Figure 11) [118].

1.2.1 Structural mass spectrometry in virology

The ultimate goal in structural virology is to understand how the structure of a virus influences its interactions with a host organism and thus governs its life cycle. Therefore, virologists nowadays employ a broad range of structural biology methods to uncover the organization of viral particles and to understand their function on a molecular level.

Macromolecular crystallography techniques, nuclear magnetic resonance spectroscopy as well as cryo-electron microscopy routinely provide three-dimensional structures [119]. Most importantly, however, all these techniques have their own challenges and limitations and there is very rarely a single technique, which could answer all questions at hand. Therefore, structural biology is moving more and more in the direction of so-called integrative approaches, where high-resolution structural data is combined with lower resolution information from different methods [120-122].

One of the experimental techniques, which provide invaluable input for the integrative structural efforts is mass spectrometry (MS). Originally belonging purely to the domains of physics and analytical chemistry, MS has changed significantly since the revolutionary developments of electrospray ionization (ESI) and matrix-assisted laser desorption-ionization (MALDI) in the late 1980s [123-125]. These two so-called 'soft-ionization techniques' instituted the analyses of biomolecules and allowed MS to expand hugely into the field of biochemistry.

Moreover, so-called structural MS has also especially established itself in the last two decades as a powerful method to study proteins. Two lines of development enabled this. On one hand, pioneering works by the groups of Carol Robinson, Albert Heck and Joseph Loo have exploited the softness of the ESI process to analyze whole non-covalently bound protein complexes without disrupting their stoichiometry and interactions in a process termed native MS [126-129]. On the other hand, many techniques were developed which probe protein conformations, interactions and dynamics directly in solution. Usually, such methods use some form of

chemical labeling, be it through isotopic exchange [130] or chemical crosslinking [131], and capitalize on the use of MS as a detection method.

1.2.2 Native mass spectrometry

Native MS employs ESI to interrogate protein structure *in vacuo* [132, 133] and closes a gap in integrative structural biology by delivering orthogonal data about all non-covalent interactions in their current state. The main challenge is retaining non-covalent interactions of proteins and their stoichiometry in complexes upon transfer into the gas phase. The results depict a dynamical landscape of protein complexes in solution.

1.2.2.1 Sample requirements

The basis of native MS is the very soft ESI [134]. The ideal protein ion is free of all solvent and buffer adducts but populated by multiple charges. Therefore, native MS presented herein requires buffer exchange into a MS-compatible solution of volatile salt. For that purpose, ammonium acetate is particularly suitable, due to its adjustable pH (4.6-10), ionic strength (0 to several M), mimicking the natural environment and providing favorable conditions for the protein until its desolvation [135]. Small amounts of non-volatile salts or small molecules, up to a few mM, can be added when required, for example bivalent cations or co-factors.

1.2.2.2 Electrospray ionization of natively folded proteins

In principle, ESI gently ionizes an analyte from a physiological solution over highly charged droplets to a charged molecular ion. By not having any size constraints, ESI enabled native MS to tackle increasingly bigger assemblies, ultimately leading to viral capsids and viral capsids [136, 137]. ESI is soft enough that intact viruses have been shown to survive the process and keep their infectivity [138, 139].

Modern mass spectrometers have a nanoelectrospray ionization (nanoESI) source, a setup spraying the sample over an electronic potential from a capillary to the entrance of the mass spectrometer, without any additional pumping devices [140]. Advantages of nanoESI are a slow flow rate and thus low sample consumption, higher sensitivity and lower background noise. To transfer charges to the sample solution, nanoESI capillaries are coated with a conductive material such as gold or simply a metal wire is inserted.

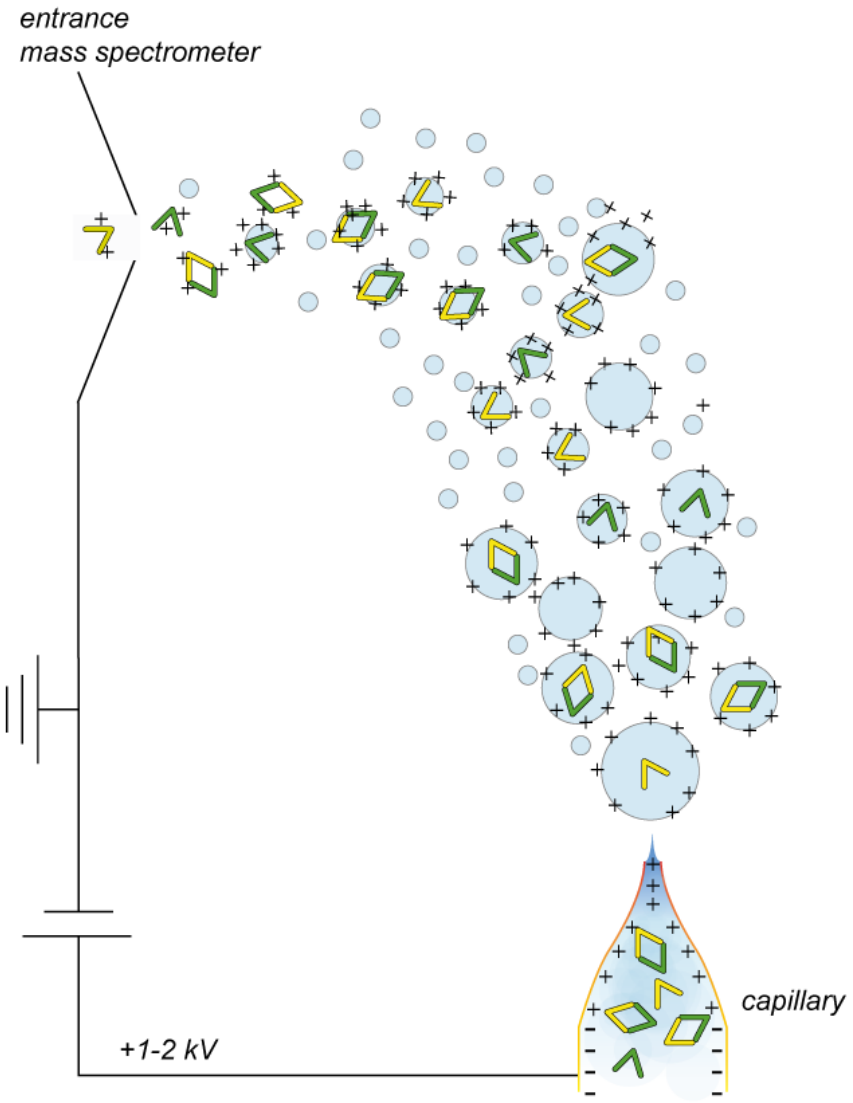


Figure 10: Schematic representation of a nanoESI source in positive ion mode. The electric potential between capillary and MS entrance induces charge separation in the bulk solution. Thereby, a spray is generated of sub-micrometer-sized droplets, containing proteins (yellow, green) and protein complexes. During the flight time, the droplets undergo shrinkage. Ideally, the proteins become charged and 'naked' molecular ions. Modified with permission from Springer Nature, BIOspektrum, "Native Massenspektrometrie für die Proteinstrukturanalytik" (Heidemann, Krichel et al.), Copyright 2018 [118].

The physical background regarding desolvation, protein charging and structural integrity of gas phase ions was unraveled experimentally and in molecular simulations [141, 142]. Ion formation starts upon inducing an electric potential between capillary and entrance of the mass spectrometer (Figure 10). Thereby, charge separation drives the solution out the capillary. At the micrometer sizes opening of the capillary, the liquid characteristically forms a Taylor cone, the place of droplet fission. Initially, the droplet size is about 150 nm, but during flight time, processes such a solvent evaporation lead to continuous shrinkage [140, 143]. As a droplet shrinks, its charges are confined on a continuously decreasing volume. The maximum amount

of charge, z_R, allowed at a given droplet radius can be predicted from the Rayleigh limit [143-145] in (1),

$$\mathbf{z_R} = \frac{\mathbf{8\pi(\gamma\varepsilon_0 R^3)^{1/2}}}{\mathbf{e}} \quad (1)$$

in which γ is the surface tension, ε_0 the electrical permittivity of vacuum, e the elementary charge and R the radius of the droplet. If the charge exceeds the Rayleigh limit, one charge ejects from the droplet by fission.

For globular proteins, the last stage of liberation from the droplet is evaporation to 'dryness', as envisioned in the charged residue model (CRM). Theoretically, if the protein remains folded, the maximum charge z_R of a protein ion is, corresponding to the final radius of the droplet, close to the Rayleigh limit. In fact, experiments confirmed that protein ions adopt charge states close to the Rayleigh limit corresponding to radii as estimated from their molecular weight in (2),

$$\mathbf{z_R = 0.078\sqrt{M}} \quad (2)$$

in which *M* is the molecular weight of the protein or protein complex in Dalton [146].

Nevertheless, analytes with structure different from a typical globular fold might have different droplet liberation mechanisms. Small ions rather leave the droplet via field emission, consistent via ion evaporation model (IEM) [147]. For unfolded or hydrophobic proteins, the chain ejection model (CEM) is a more accurate description of droplet liberation. Before complete evaporation, the analyte remains close to the liquid-air interface but gradually unfolds out of the droplet. The unfolding chain adopts migrating charges, which results in a highly charged ion ejected from the droplet [148].

In native MS spectra, a signal from a single ion species typically appears in a Gaussian distributed envelope of peaks (Figure 11). In positive ion mode, the multiple peaks represent multiple molecular ions of one single mass species, M, populated by varying amount of protons, H^+, and therefore charges, z. The molecular ion $[M+zH]^{z+}$ determines the value of the m/z signal $\left(\frac{m}{z}\right)$. Assuming that the mass of the proton is $m_H = 1$ the molecular ion is treated as $[M+z]^z$. With these theoretical ions the number of charges, z, and the molecular weight, M, from adjacent peak signals, $\left(\frac{m_1}{z_1}\right)$ and $\left(\frac{m_2}{z_2}\right)$ can be easily calculated [149], which is briefly exemplified on the next page.

When both signals derive from the mass species, M, the observed mass-to-charge ratios follow from (3), (4), and (5):

$$\mathbf{z_1 = z_2 + 1} \tag{3}$$

$$\left(\frac{\mathbf{m_2}}{\mathbf{z_2}}\right) = \frac{\mathbf{M} + \mathbf{z_2}}{\mathbf{z_2}} \tag{4}$$

$$\left(\frac{\mathbf{m_1}}{\mathbf{z_1}}\right) = \frac{\mathbf{M} + (\mathbf{z_2 + 1})}{\mathbf{z_2 + 1}} \tag{5}$$

For obtaining the charge state z_2, equations (4) and (5) are rearranged for M in (6) and (7) and substituted in (8):

$$\mathbf{M = z_2}\left(\left(\frac{\mathbf{m_2}}{\mathbf{z_2}}\right) - \mathbf{1}\right) \tag{6}$$

$$\mathbf{M = z_2}\left(\left(\frac{\mathbf{m_1}}{\mathbf{z_1}}\right) - \mathbf{1}\right) + \left(\frac{\mathbf{m_1}}{\mathbf{z_1}}\right) + 1 \tag{7}$$

$$\mathbf{z_2} = \frac{\left(\frac{\mathbf{m_1}}{\mathbf{z_1}}\right) - \mathbf{1}}{\left(\frac{\mathbf{m_2}}{\mathbf{z_2}}\right) - \left(\frac{\mathbf{m_1}}{\mathbf{z_1}}\right)} \tag{8}$$

For obtaining the mass M of the deduced charge z_2 from (6) is inserted in (9):

$$\mathbf{M = z_2}\left(\left(\frac{\mathbf{m_2}}{\mathbf{z_2}}\right) - \mathbf{1}\right) \tag{9}$$

This method is useful when quickly determining z and M from two adjacent peaks. However, considering merely two peaks leads to a larger error. Therefore, in routinely measurements the masses are determined computationally by MS software such as Masslynx (Waters). Therein, peak signals require smoothing and centring before the mass of an analyte is calculated from a peak series, considering all peaks in one series.

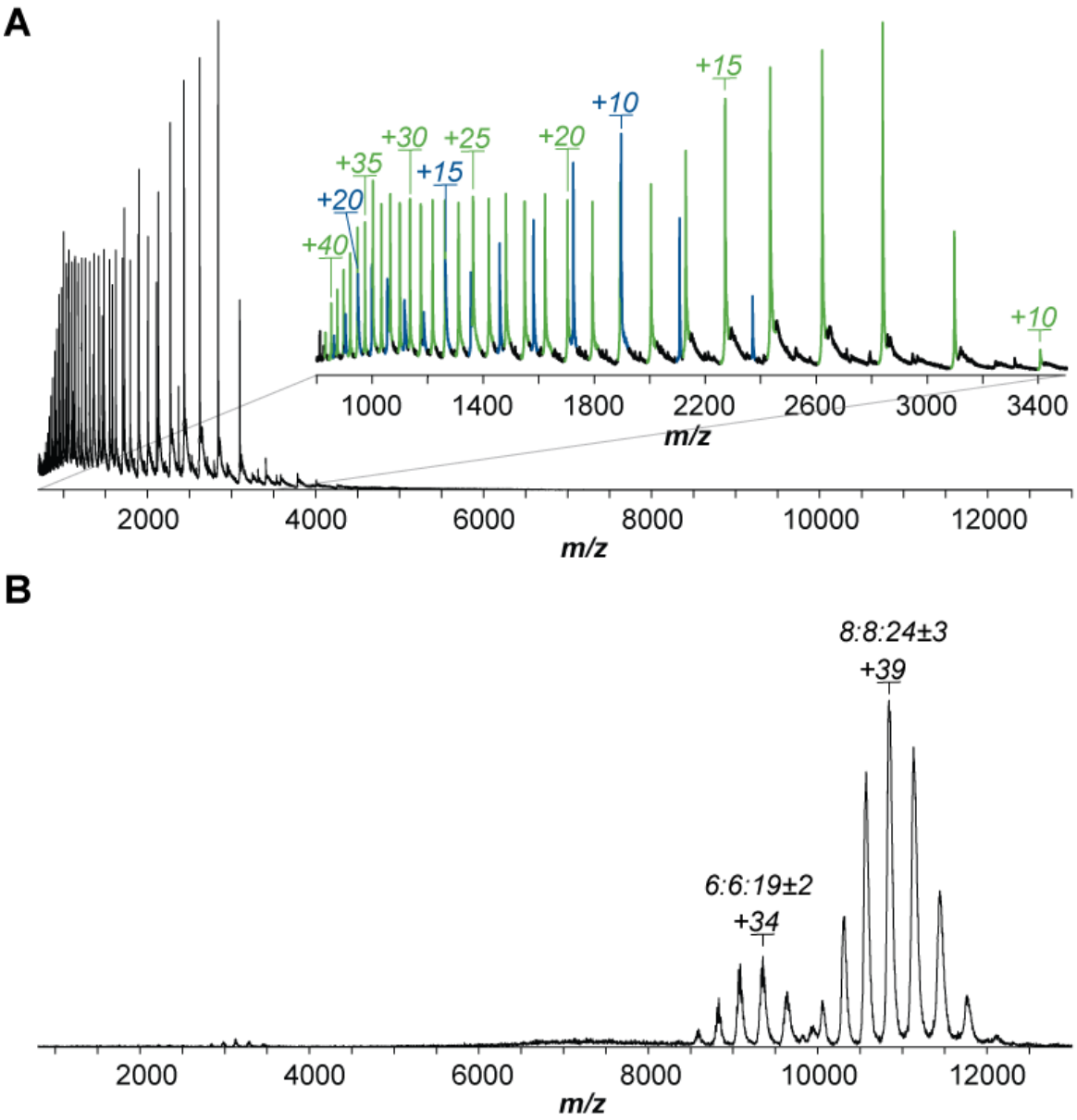


Figure 11: Denaturing and native MS conditions of a protein-lipid complex. (A) Denaturing ESI conditions in 0.5 % formic acid epsin-2 (blue) and Sla2 (green) resulting in two independent peak series in the lower *m*/*z*-range. The high charge states indicating for unfolding but allowing a precise analysis of the intact mass (epsin-2: 18.958 ± 4 Da, Sla2: 30.408 ± 10 Da. **(B)** Native ESI conditions of the same proteins in 300 mM AmAc pH8 in presence of the phospholipid phosphatidylinositol-4,5-bisphosphate (PI(4,5)P2) revealed two different complexes in the high *m*/*z*-range with the indicated epsin-2:Sla2:PI(4,5)P2 stoichiometry. The observed complexes indicating retained structural integrity in the gas phase. The native MS conditions mimicking the physiological solution combined with ESI allow for investigation of the protein complex in the gas phase. Modified with permission from Springer Nature, BIOspektrum, "Native Massenspektrometrie für die Proteinstrukturanalytik" [118], Copyright 2018.

1.2.3 Protein structure in the gas-phase

To retain protein structure in solution, the hydrophobic effect is a key stabilizing factor [150], but in gas phase this effect is absent. Simplified, this means that in the gas phase, hydrophobic interactions are weakened or non-existing and polar interactions are strengthened. To which degree protein structures in solution resemble protein structures in the gas phase is ongoing matter of investigation. Even though most studies also address side-chain collapse, unfolding and refolding into new non-native structures, they find evidence that other structural aspects are retained [151-153].

The charge states of protein ions correspond with the laid out model of CRM, suggesting that proteins remain compact upon electrospray ionization when in buffer surrogates mimicking physiological conditions (e.g. 150 mM AmAc, pH 7.5) [154-156]. The collisional cross-section (CCS^2) in the gas phase determined by IM-MS shows that higher charged ions have an extended conformation, probably due to unfolding and electrostatic repulsion, while lower charged ions are more compact. Their gas phase CCS^2 correspond to CCS^2 from X-ray crystallography [157, 158], but is slightly smaller, probably due to the collapsing side chains and absence of water molecules. Adding temporal resolution, Wyttenbach et al. reported that ubiquitin retains its solution structure in the gas phase for more than 100 ms, longer than typical native MS measurements [157]. Moreover, infrared spectroscopy (IR) found that the protein secondary structures α-helices and β-sheets are retained during flight [159].

All together, it is widely accepted that gas phase conformers of ions often represent metastable, kinetically trapped species, which are 'frozen out' as a result of evaporative cooling as ions are formed [153, 160]. Therein, activation barriers prevent large-scale transitions such as unfolding and central aspects of the structures remain stable for the time scale of analysis.

1.2.4 Mass analyzers

The heart of a mass spectrometer is a device separating a mixture of gas-phase ions according to mass and charge and their relative abundances. In recent years, native MS has progressed towards higher resolution mass analyzers by the introduction of the high-mass capable Orbitrap (orbital ion trap), which complements the current time-of-flight (TOF) devices. However, while Orbitraps have higher mass resolution, TOF platforms are used in parallel because they offer a higher dynamic range, speed and sensitivity [161].

In order to perform collision-induced-dissociation (CID) of gas phase ions, quadrupole (Q) – Time-Of-Flight (TOF) mass spectrometers are popular. These instruments have two interconnected mass analyzers for different purposes, a Q to use as a low-resolution mass filter and a TOF for a high-resolution readout. Since a Q-TOF was used conducting the main experiments in this thesis, both mass analyzers are described in detail.

1.2.4.1 Time-of-Flight

The principle of TOF is based on gas phase ions simultaneously starting a journey through a flight tube to the detector where their flight time is determined, which corresponds to their *m*/*z*. An unprecedented dynamic range is reached measuring the flight time of all gas phase ions from a sample mixture in one working circle. Therefore, TOF analyzers are routinely used for native MS, allowing measurements of large protein complexes and their non-interacting subunits in one go [161].

While TOF mass analyzers exist in different setups, they are all based on the same physical principle of flight time determination. Upon injection in the flight tube, the ions potential energy is converted into kinetic energy as follows from (10) and (11),

$$\mathbf{E_{kin}} = \mathbf{E_{pot}} \tag{10}$$

$$\mathbf{zeV} = \frac{\mathbf{1}}{\mathbf{2}}\mathbf{mv^2} \tag{11}$$

in which *z* is the number of elementary charges e, V is the applied acceleration voltage, and m and v are the mass and velocity of the ion, respectively.

The velocity is only dependent on flight time t and length of trajectory in the flight tube d (12).

$$\frac{\mathbf{d}}{\mathbf{t}} = \mathbf{v} \tag{12}$$

Substituting (12)) into (11) leads to (13), which is rearranged to express the time (14):

$$\mathbf{zeV} = \frac{\mathbf{1}}{\mathbf{2}}\mathbf{m}\left(\frac{\mathbf{d}}{\mathbf{t}}\right)^{\mathbf{2}} \tag{13}$$

$$\mathbf{t} = \frac{\mathbf{d}}{\sqrt{\mathbf{2eV}}}\sqrt{\frac{\mathbf{m}}{\mathbf{z}}} \tag{14}$$

Usually, d, e and V are instrument constants or adjustable parameters, and therefore the flight time is only dependent on $\sqrt{\frac{m}{z}}$ of the analyte ion. This means that ions of lower mass or higher charge fly faster. Later, *m*/*z* is used as the x-axis of a typical mass spectrum.

Historically, TOF analyzers were established in the 1960s, but their final success took hold after two specific instrument developments, orthogonal injection and the reflectron. Orthogonal injection makes use of a pusher device that sends ion packages from their linear trajectory into a flight tube, resulting in a similar starting time [162]. However, injected ions, otherwise similar, have deviating initial energies that originate from different vertical positions, and therefore

impose a limit for ion separation. The reflectron is an ion mirror that compensates for the initial energy and spatial spread and thereby, tremendously increases resolving power [163].

At the end of the trajectory, the ions flight time finishes upon impact on a detector, typically a multi-channel plate (MCP). These detectors consist of an array of charged electron multipliers. Ion impact close to a channel entrance strikes a burst of electrons. The microchannel leads the electrons to a time-to-digital converter, which records the electric signal. Due to the short path width of the channels, which consequently have a short pulse length and high temporal resolution (< ns), these devices became the detector of choice for TOF-MS [164].

1.2.4.2 Quadrupole

Quadrupole mass analyzers consist of four parallel rod electrodes arranged as square whose opposing electrode pairs are connected. The electric potential between the connected rods is a superposition of a static and sinusoidal radio frequency (RF) potential resulting in motion of ions along the rods.

In native MS instruments modified for high-mass measurements, quadrupoles operate at an lower frequency that sets the transmission limit higher and therefore enables native MS of large protein complexes [165]. Nevertheless, in these specific instruments, quadrupoles are mainly applied as mass filters due to the ability of a set RF amplitude and frequency to allow only a specific range of m/z ions to pass [162, 166]. Thereby, these mass spectrometers exploit the quadrupole as a narrow-band m/z filter to separate selected ions for subsequent gas phase analysis.

1.2.5 Collision-induced dissociation

Determining the mass of a protein complex is not always sufficient. To assign stoichiometry unambiguously, an ion of the complex, termed precursor ion, is selected in a quadrupole and dissociated in the gas phase (Native CID-MS workflow: Figure 12).

For this purpose, collision induced dissociation (CID) is mostly used, in which additional energy is imparted to the precursor via high-velocity collisions with inert gas atoms. This gradually increases the internal energy of the ion until individual subunits of a complex unfold, lose their non-covalent interactions and are eventually ejected from the complex one-by-one [167]. Mass information obtained from the ejected subunits and the remaining subcomplexes proves stoichiometry of the originating complex.

The order of subunit ejection also correlates with its peripheral location within the complex and thus provides even more insight into the complex architecture. When harsh CID conditions are used, backbone fragmentation of the intact protein occurs on top of subunit dissociation. Notably, this dissociation behavior in the gas phase is usually distinct from solution, where dissociation is a spontaneous and reversible equilibrium reaction rather than a consequence of physical collisions with gas molecules [168].

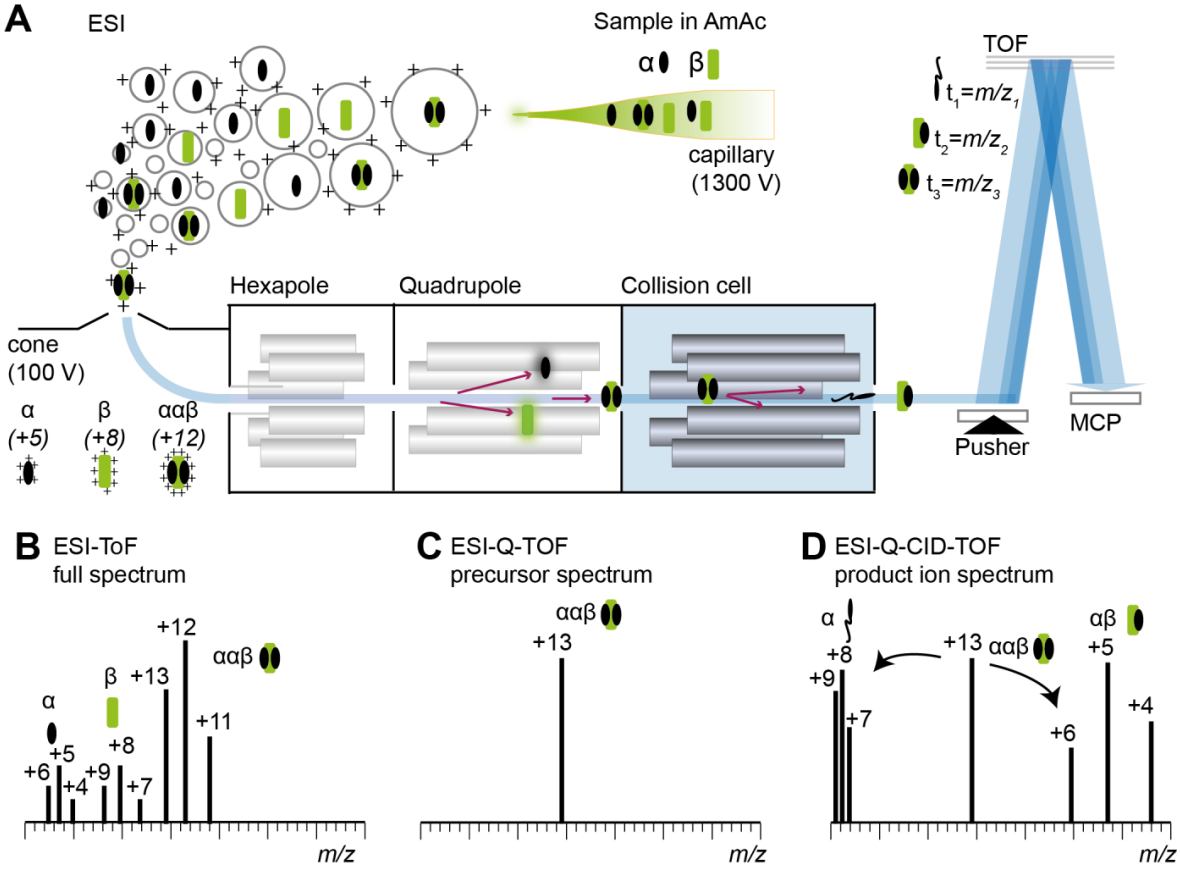


Figure 12: Native MS TOF-MS and tandem MS workflow and readout. (A) Protein complex with two subunits (black and green) sprayed from ammonium acetate solution over an electric potential and enters the mass spectrometer as molecular ion. The modules of the mass spectrometer allow for manipulation and analysis *in vacuo* as depicted in schematic spectra. **(B)** Full ESI-MS spectrum reveals a Gaussian distribution of the charge states of each mass of the sample proteins. An additional charge distribution indicates a trimeric complex. **(C)** To confirm subunit stoichiometry of the candidate complex, first precursor ion spectra are acquired to control filtering for a molecular ion in the quadrupole region of the mass spectrometer. **(D)** Product ion spectra from collision of precursor ion with a neutral gas in the collision cell. The complex dissociates into two characteristic products confirming (2:1) stoichiometry. The dissociated subunit unfolds withdrawing charges from the remaining ion. Therefore, the unfolded subunit is detected in the lower *m/z* range of the spectrum, whereas the remainder of the complex is detected at higher m/z. This figure is modified from Dülfer, Kadek, Kopicki, Krichel and Uetrecht, under submission (2019).

1.2.6 Q-TOF2 modified for high masses

The mass spectrometer used for the experiments in this work is a high-mass Q-TOF. The main successively interconnected modules characterizing this mass spectrometer are the nanoESI, hexapole, quadrupole, collision cell and Time-of-Flight.

Originally, the Q-TOF had been introduced for analysis of smaller ions such as metabolites or peptides, but in 2006, van den Heuvel and van Duijn et al. presented this instrument with a set of adjustments, which were essential in order to perform native MS of relatively large analytes such as protein complexes [165]. The key modification for transmission of high mass ions is the elevated pressure in the source regions hexapole, which focusses ions for their further trajectory. Normally, due to the high energy of larger ions, they would not be able to hold the path in a vacuum. To reduce the ions energy, a flow-restricting sleeve is set around the first hexapole ion bridge, resulting in an elevated pressure. By colliding with ambient gas, the ions undergo an energy-reducing process termed collisional focussing and have a higher chance of a stable trajectory.

A further modification is to improve CID precursor selection and transmission of larger ions by reducing RF frequency of the quadrupole. Upon these changes, the quadrupole selects ions with m/z up to 32000 and transmits ions in broad-band mode up to 100000 m/z.

Furthermore, CID of larger protein complexes is enabled by a high-pressure hexapole collision cell, which is operated at approx. 1.5×10^{-2} mbar with gases such as argon or xenon. Additional modifications to improve transmission of larger ions are high-transmission grids in the multicomponent ion lenses and a low repetition rate pusher.

1.2.7 Quantitative native MS

A desired feature is to deduce protein concentrations from signal intensities in mass spectra. In general, this is possible because in ESI-MS the signal response, I_x, is proportional to protein concentration, c_x, when the ionization efficiency r_x is known (15).

$$\mathbf{I_x = r_x * c_x} \tag{15}$$

The r_x basically describes the fraction of a species in solution being ionized and resulting in a signal at the detector [169]. This parameter depends on multiple variables such as analyte ion efficiency and instrumental bias. Therefore, the response factor r_x for each species x has to be determined experimentally in titration measurements.

For two analytes, the signal intensity is proportional to the analytes concentration. However, especially large oligomer complexes deviate from a direct correlation at higher concentrations which hampers reliable quantification [170]. For a Q-TOF instrument as used in this work, 20 μM appears to be an upper limit for proportional signal intensity due to ion suppression between proteins.

Even though theoretically possible, signal responses of protein mixtures have not yet been directly determined. However, for heterogeneous aggregates of a single protein it was shown that signal intensity over a broad range of a mass spectrum, multiplied for n-oligomeric complexes, correlates well with the mass fractions as determined by absorption and size exclusion chromatography. These measurements showed that instrument bias and ion efficiency is in fact not overall decisive and that ESI-TOF mass spectra could be analyzed in a semi-quantitative fashion [171, 172] by calculating the mass fractions.

Mixtures of different proteins would become complicated, because each analyte (x_1, x_2…x_n) has their own specific signal response (r_{x1}, r_{x2} … r_{xn}). Moreover, the species within one mixture might have been influenced during sample preparation or suppress each other's ion efficiency. In dynamical systems of multiprotein complexes, the response factors are probably dynamic as well, and therefore difficult if not impossible to determine. Nevertheless, increasing or decreasing relative intensities of species in a mixture indicate a relative change in concentration.

1.2.8 Hydrogen-deuterium exchange mass spectrometry

Hydrogen-deuterium exchange (HDX) MS is a sensitive technique to localize regions of conformational dynamics in proteins and identify amino acids engaged in intermolecular interaction under close to native conditions in solution. As opposed to other MS experiments, HDX can be performed in practically any buffer condition and with additives or co-factors that are required for proper protein folding and function – the sole difference is buffer preparation from deuterated water. The method exploits the naturally occurring exchange between solvent-accessible protein backbone hydrogens and deuterium in solution over time, measured as mass increase using MS [173, 174].

The most common – bottom-up continuous labeling - HDX experiments compare two proteins in different states (HDX-Workflow: Figure 13). The studied protein is labeled in a deuterated buffer under native conditions. After various time intervals, the labeling reaction is quenched by rapidly reducing the pH to 2.5 and the temperature to 0°C, where the HDX rate and therefore also the rate of D/H back exchange, is minimal. Under these conditions, the half-life of the deuterium label (throughout subsequent LC-MS analysis performed in non-deuterated solvents) is sufficiently long [175]. To localize deuterium incorporation, the protein is split up into smaller peptides prior to LC-MS analysis. These peptides can be generated via digestion with an aspartic protease, either in bulk solution or more conveniently using protease immobilized on a flow-through column [176]. Due to the pH required to keep deuterium back-exchange loss minimal, porcine pepsin is routinely used as the protease of choice for HDX-MS applications. However, a number of alternatives are used, which can yield different peptides and more precise deuteration localization [177-180].

Following digestion, peptides are separated in a liquid chromatography (LC) system and directly electrosprayed into a mass spectrometer. There, created peptides are first identified by their mass-to-charge ratio (m/z) and MS/MS fragmentation patterns from non-deuterated samples. Subsequently, their deuterium content is measured in the samples exposed to deuterium and compared between experimental conditions to localize any differences.

The rate of exchange is influenced by several factors, including solution conditions such as pH and temperature as well as by the peptide's local environment [181]. Most importantly, from the point of view of protein structure and interactions, the involvement of hydrogens in hydrogen bonding within higher order elements of protein structure (e.g. in α-helices and β-sheets) or in interactions with ligands slows down the exchange rate for respective backbone amide hydrogens. Further, protection from solvent, for instance in a hydrophobic core of the protein or at a protein-protein interface, also typically leads to a decrease in exchange in the affected protein region. Additionally, HDX can also detect more distant effects induced by an

interaction, for example through allostery [182]. Finally, conformational dynamics and flexibility of protein regions play a role in influencing the exchange rate too, as the exposure of regions to the surrounding deuterated solvent accelerates the HDX. This can sometimes even be reflected in more complex exchange kinetics, which uncover additional information on distinct conformational dynamics [183].

Considering the range of monitored changes within proteins, HDX-MS is a versatile tool to gain insights into the dynamics and structural responses of protein backbones to perturbed solution conditions and ligand or protein binding [184, 185].

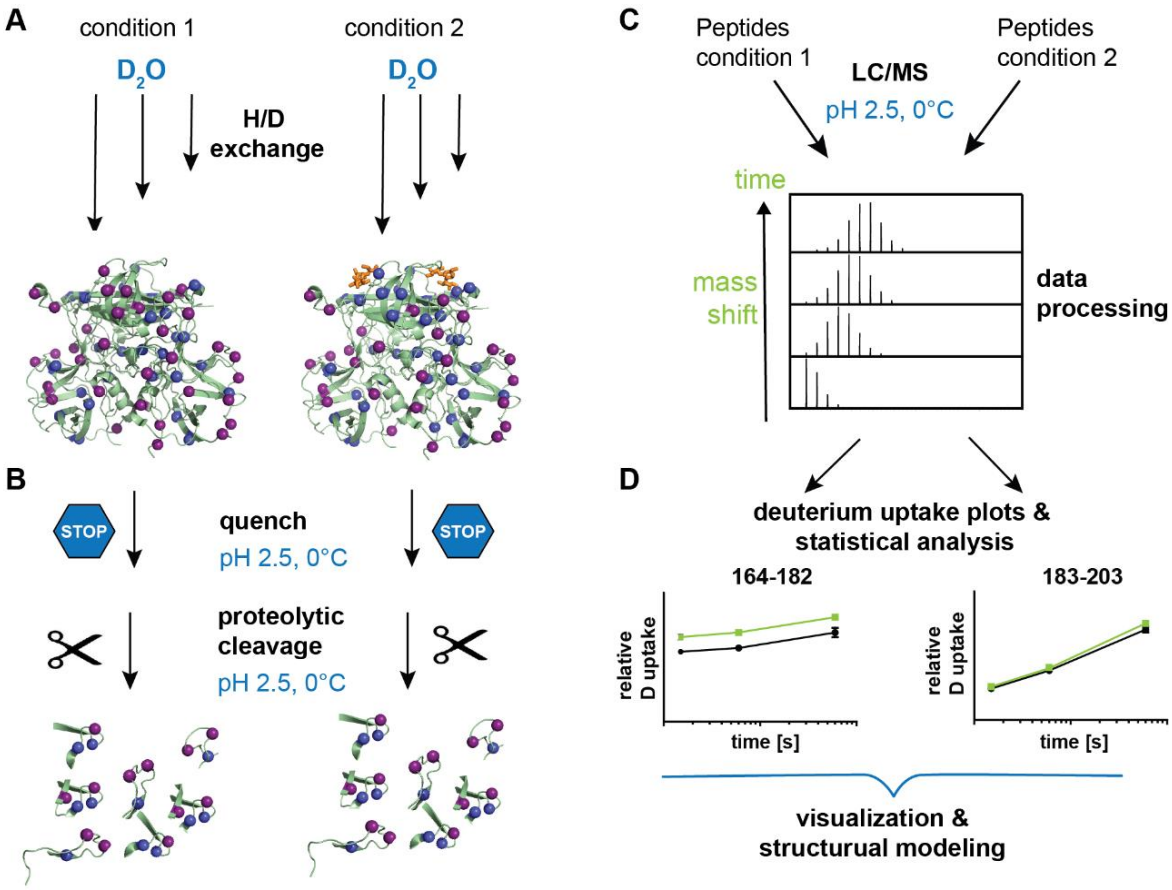


Figure 13: Hydrogen-deuterium exchange workflow. Exemplarily workflow for a continuous labeling experiment comparing two conditions of a protein. **(A)** Samples are prepared in two different conditions, for example one protein with and without small molecule binding partner (orange). In the H/D exchange step, both samples are labelled in deuterated water for several time intervals. **(B)** Then, a quenching step reduces HDX to a minimum and proteolytic cleavage results in peptides. **(C)** These peptides are separated and analyzed by LC-MS. The observed mass shifts indicate for different deuterium uptake between time-points and samples. **(D)** Deuterium uptake plots showing differences (green vs. black curve) for peptides aa164-182 und aa183-203 under the analyzed conditions, for example here small molecule binding. For interpretation of the acquired data, structural models are refined or deuterium uptake differences visualized on existing structural models.

1.2.9 MALDI mass spectrometry of cross-linked protein complexes

Matrix-assisted laser desorption ionization (MALDI) is considered, like ESI, a soft ionization method. The typical MALDI methodology has its strength in analysis of smaller biomolecules, but when adjusted can acquire the mass spectra of larger proteins and protein complexes with advantages over native MS [186-188] (**Figure** 14).

The principle of MALDI is based on ionization of a matrix-sample mix and enables a high throughput analysis that tolerates salts and detergents. The properties of the matrix is critical for successful ionization. Most importantly, the matrix needs to dissolve the analyte, co-crystallize, co-desorb and show reactivity to the laser wavelength [189]. While many matrices are available, for the analysis of proteins sinapinic acid (SA) is well suited to generate singly charge ions form larger proteins.

The mechanistic behaviour of particles in a drying droplet are explained in the coffee-ring effect, whose name comes from the typical ring shaped stain observed upon spilling a drop of coffee [190]. It describes how dissolved colloidal and globular shaped particles migrate to the sites of the droplet edge, drying there early, and forming a ring. Elongated particles decrease the effect by rather remaining on the surface of a drying droplet. The coffee ring effect has certain implications for the co-crystallization of high-mass biomolecules [191]. In the so-called first-shot phenomenon, MALDI ionization of larger proteins yield higher from the exterior of a spot [192, 193]. Most probably, larger proteins co-crystallize in a size-exclusion-chromatography-like process and therefore locate preferably on the sites or surface on a MALDI spot, while smaller proteins remain evenly distributed.

High mass MALDI is often performed in linear TOF mode, resulting in relatively low resolution of the mass spectra. Moreover, MCP detectors do not generate efficient electronic signal from low energy ions and limits detection over 100,000 *m*/*z*. Therefore, high-mass detectors must be installed, such as conversion dynodes, enabling detection of more than 1,500,000 m/z [194].

Some researchers analysed protein complexes with MALDI-MS, exploiting the first-shot effect and matrices at near neutral pH [193, 195, 196]. However, the requirements of an acidified matrix and sufficient laser energy for effective ionization disturb non-covalent interactions. Therefore, it has become a common methodology to stabilize non-covalent interactions by using chemical cross-linkers, which survive the ionization process [187]. Different from ESI, where volatile substances are required, cross-linking is possible in a physiological buffer where complexes have a higher probability to assemble correctly and therefore observed interactions could have more relevance.

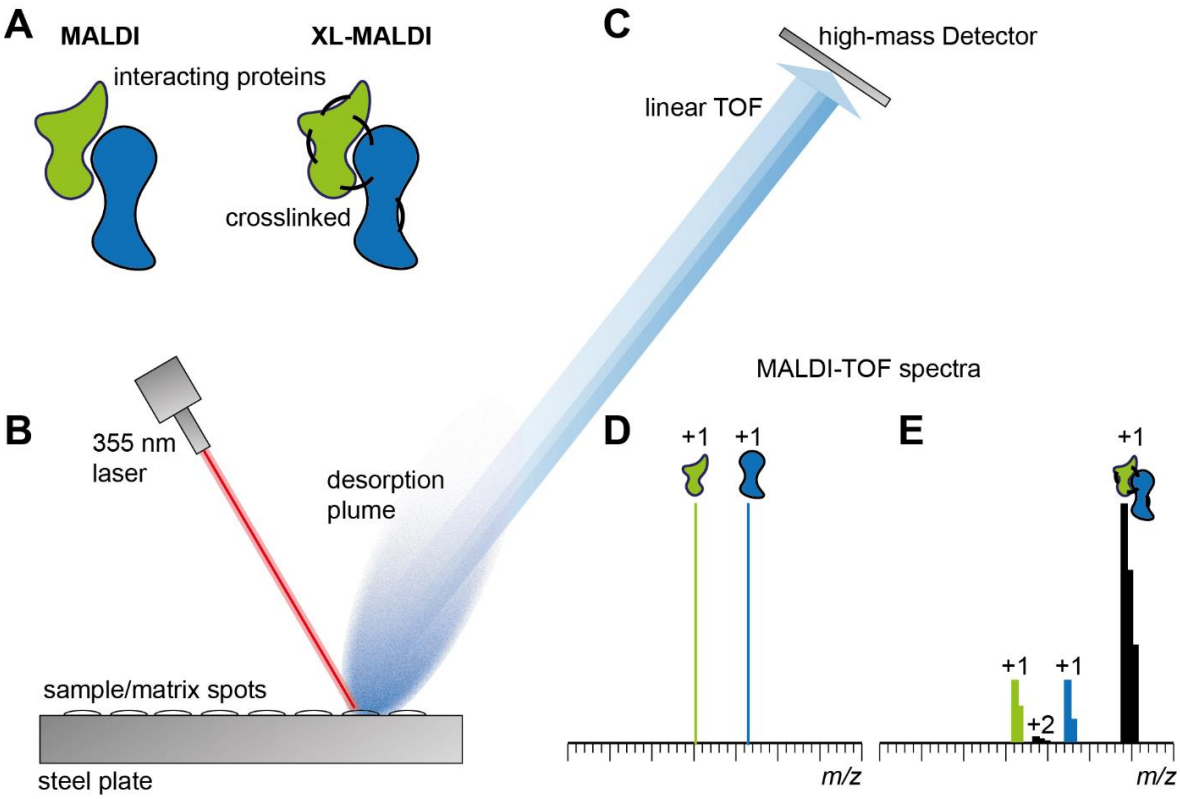


Figure 14: Cross-linking coupled to high-mass-MALDI-MS workflow. (A) Hetero-dimeric protein complex (green-blue) stabilized with a cross-linker. **(B)** First sample/matrix mixture is spotted on a steel plate and allowed to co-crystallize before a pulsed laser induces plume from the spots. **(C)** Subsequently, the analyte ions are separated by their *m*/*z* ratio in a linear TOF. **(D)** MALDI-TOF mass spectrum for the protein complex showing two signals for singly charged ions of each subunit. **(E)** MALDI -TOF mass spectrum of cross-linked sample showing additionally signals for the complex, singly and doubly charged. Peaks mass shifted and broadened due to the effect of attached cross linker.

The combination of cross-linking (XL) and MS is a vivid field within in structural MS [197]. Cross-linkers are widely available and constantly improve properties, for instance length of the linker, specificity with mono and bi-functionality or photo-activation. Consequently, the cross-linker has to be suited to the research question. One kind of cross-linkers often used for high-mass MALDI are NHS-esters tackling primary amines in the amino acid lysine and the proteins N-terminus [198, 199].

Furthermore, a 'magic pill' for cross-linking is the super-reactive glutaraldehyde, tackling protein amines, thiols, phenols, and imidazoles [200]. The exact structure and reaction mechanism of glutaraldehyde is still matter of ongoing research. Nevertheless, the molecule is infamous for its high reactivity, leading to unspecific crosslinks or resulting in aggregation of the complete sample. Many 'mono-links' increase the observed mass and shift the signals to a higher *m*/*z*. However, these issues could be attributed to any cross-linker. Therefore, crosslinking experiments require careful adjustments to find the sweet spot, optimizing concentration of cross-linker and protein as well as reaction time.

2 Aim and objective

Aim of this thesis was to characterize structural changes and dynamics upon processing of coronavirus NSP7-10 by means of structural mass spectrometry.

Processing in the beginning of infection could have regulatory implications for a functional polyprotein, *cis*-acting protein complex, or binding scaffold for core enzymes. Processing is timely coordinated by the specific activity of M_{pro} at the cleavage sites. These have been studied with peptides and conserved cleavage efficiencies could have implications for cleavage order. However, peptide substrates display only a small stretch of amino acids of a cleavage site. Full-length proteins consider more structural aspects from a folded protein and display a native cleavage site. It is not known, if intermediate products and final products that are processed from full-length proteins correspond to conserved relative efficiencies found at peptides.

Binding of NSP7+8 and NSP8 independently at two different sites of the main polymerase, is one key interaction of the coronavirus replication-transcription complex. X-ray crystallography structures of two full-length NSP7+8 complexes have been reported and show stoichiometric differences between species. This raised the questions of do more structures of NSP7+8 exist and which structural features influence stoichiometry.

To prepare the experiments, full-length NSP protein and M_{pro} samples will be produced recombinantly in *E.coli* and characterized in native MS.

In a first set of experiments, the processing of different CoV species full-length protein substrates will be investigated with native mass spectrometry in order to reveal relative cleavage efficiencies and the dynamical change of mass species from substrate over intermediate products to final products. Additionally, a FRET peptide assay will be preformed to find out if both substrates are equally suited and if proposed conservation of cleavage site efficiency holds true.

In a second set of experiments, complex formation will be studied of NSP7+8 from different species of various degrees of sequence conservation. Native MS will picture the landscape of non-covalent interactions and collision-induced dissociation in Q-TOF tandem MS will hint on architectural differences between different complexes.

In a third set of experiments, a high mass and cross-linking approach will be exploited in MALDI-MS to get a different view on processing and complex formation.

Finally, in a fourth set of experiments, a bottom-up HDX-MS approach will be tested to reveal interaction sites and structural dynamics upon cleavage and rearrangement of a NSP7-8 precursor protein to a NSP7+8 protein complex.

3 Materials and methods

3.1 Cloning and gene constructs

The plasmids used in this work were either cloned or re-cloned in house, directly acquired from collaborators or not handled since purified protein was obtained from collaborators. Details of expression plasmids that are relevant for this work are listed below (Table 3) and the resulting masses of the protein products are listed in the supplement (Table S 1: Molecular weights (MWs) of NSPs, their cleavage products and complexes.).

To generate inserts for expression plasmids for CoVs ORF1ab NSP6-10 region, DNA was amplified by PCR from commercially applied cDNA (Eurofins scientific SE) and from DNA plasmids existing from earlier work [83]. To create ends for directed cloning the PCR products were digested by Eco31I (BsaI) (Thermo Fisher Scientific) and ligated into IBA pASK33+ and pASK35+, encoding for C- and N-terminal His-tag, respectively.

The expression plasmid for SARS-CoV M_{pro} with authentic ends was obtained from collaborators, and had been created as described by Xue and Yang et al.[56].

Table 3: Expression plasmids for recombinant proteins. Amino acids in one letter code refer to all residues between tag and NSP sequence of interest. Sources of plasmids and proteins were Sven Falke (SF), Rolf Hilgenfeld (RH), Lars Redecke (LR), Ganesh Bylapudi (GB) and John Ziebuhr (JZ). Abbreviations for virus species were shortened to simplify the protein names e.g from SARS-CoV to SARS or from hCoV-229e to 229e.

Name	Plasmid name	tag and linker	cleavable by M_{pro}	source of material
M_{pro} auth. end.	pGEX-6p-1/pGSTM	GST-AVLQ… …GP-6xHis	yes, no	LR,RH (plasmid) original [56]
229e NSP7-10	pRSET-A	6x-His-G…	no	LR,SF (plasmid)
SARS NSP7-10	pASK-IBA35plus	6xHis-G…	no	LR,SF (plasmid)
SARS NSP7-9	pASK-IBA33plus	…ASRGS-6xHis	yes	LR,SF (plasmid)
SARS NSP7-8	pASK-IBA33plus	…ASRGS-6xHis	yes	LR,SF (plasmid)
SARS NSP7-8	pASK-IBA35plus	6xHis-G-…	no	cloned in house
SARS NSP8-9 QAtag	pASK-IBA33plus	…ASRGS-6xHis	no	cloned in house
SARS NSP9-10 QAtag	pASK-IBA33plus	…ASRGS-6xHis	no	cloned in house
SARS NSP9	pASK-IBA35plus	6xHis-AVLQ…	yes	cloned in house
229e NSP9	pASK-IBA35plus	6xHis-AVLQ…	yes	cloned in house
229e and PEDV NSP7-8	-	…SGSG-6xHis	yes	JZ, GB (protein)
FIP and TGEV NSP7-8	-	…NNE-6xHis	yes	JZ, GB (protein)
FIP NSP7-9	-	…-6xHis	no	JZ, GB (protein)

3.2 Expression and purification

Main protease SARS-CoV M_{pro} was produced with authentic ends [56] .Briefly, the protein constructs had a GST tag connected via an auto-cleavage site at the N-terminus and a His-tag connected via a Pre-Scission protease cleavage site at the C-terminus. For expression, Bl21 Rosetta2 (Merck Millipore) were transformed with the plasmid The cells were grown to 0.4-0.6

OD_{600}, then induced with 0.5 mM IPTG and continued to grown at 20°C overnight. All further steps were carried out at 4°C. To separate soluble protein, cells were lysed in buffer T1 (40 mM Tris, 300 mM NaCl, pH 7.5) with one freeze-thaw cycle followed by sonication (Micro tip, 70 % power, 6 times on 10 s, off 60 s; Branson digital sonifier SFX 150) and centrifugation (20,000 g for 45 min). The protease was purified by affinity chromatography with Ni^{2+}-NTA beads (Thermo Fisher Scientific) and dialysis into storage buffer (50 mM HEPES, 10 % Glycerol pH7.5). Alternatively, to enhance M_{pro} activity, the His-tag was cleaved by dialysis into PreScission cleavage buffer (20 mM Tris-HCl, 150 mM NaCl, 1mM EDTA, 1 mM DTT at pH7) while incubating with PreScission protease (protease to protein ratio 1:2000) (GE Healthcare) overnight. PreScission protease was removed by binding with GST-Sepharose (GE Healthcare) for 2 h and buffer exchange with centrifugal filter devices (10,000 MWCO, Amicon, Merck Millipore) into storage buffer. Purified M_{pro} was frozen in liquid nitrogen and stored at -80°C.

To produce proteins of different-length of NSP6-10 region Bl21 Rosetta2 (Merck Millipore) were transformed with corresponding plasmids, grown in culture flasks to 0.4-0.6 OD_{600} in 0.3-1 L and then induced with 50 μM anhydrotetracycline and continued to grown at 20°C overnight. To pellet, cultures were centrifuged (6000 g for 20 min) discarded of medium and frozen at -20°C. All further steps were carried out at 4°C. To separate soluble NSPs, pelleted cells were lysed in 1:5 (v/v) buffer B1 (40 mM phosphate buffer, 300 mM NaCl) with one freeze-thaw cycle, sonicated (micro tip, 70 % power, 6 times on 10 s, off 60 s; Branson digital sonifier SFX 150) and then centrifuged at 20,000 g for 45 minutes. Proteins were isolated with Ni^{2+}-NTA beads (Thermo Fisher Scientific) in gravity flow columns (BioRad). First the beads were equilibrated with 20 column volumes (CV) B1 + 20 mM imidazole, then bound to NSPs by incubation with crude extract for 60 minutes and finally to remove unspecifically bound proteins, washed with 20 CV B1 +20 mM imidazole followed by 10 CV of B1 +50 mM imidazole. To elute the NSPs, eight fractions of 0.5 CV B1 +300 mM imidazole were collected. Immediately after elution, the fractions were spiked with 4 mM DTT.

Table 4: Most relevant molar absorption coefficients at 280 nm. The specific absorption coefficient ε was calculated from the theoretical amino acid sequence with reduced cysteine with the software *Protein calculator v3.4* (The Scripps).

Protein	ε (mol^{-1} cm^{-1})
M_{pro}	31150
SARS NSP7-8	31304
SARS NSP7-10	58513
229e NSP7-10	55900

The protein concentration of elution fractions was checked by UV absorption at 280 nm (A_{280}) (DeNovix DS11-FX+) before pooling and concentrating to 5-10 absorption units (AU). Size exclusion chromatography was performed by running 500 μL of sample volume in a

Superdex200 10/300 (GE Healthcare) column. Fractions from the center of the main peak were collected for further experiments and their concentration was checked with the specific absorption coefficient ε at A_{280} (Table 4).

The NSPs received from our collaborators (Table 3) were produced in *E. coli* TB1 cells and purified by metal-ion affinity, anion exchange and size-exclusion chromatography.

3.3 SDS-PAGE

For quality control of expression and purification SDS-PAGE was performed. Gel samples were prepared by mixing with 4x sample buffer (Table 5) in the ratio of 5:1, 1:2 and 1:4 for pellet, crude extract/flow through and elution fractions, respectively. The gel pockets were filled with samples of equal volumes (5-15 µL), depending on the analyzed sample. Gels were manually produced in a multiple gel caster (Hoefer) with 4 % collection gel and 15% separation gel (Table 6) and run in a mini gel chamber (Hoefer) at constant 25 mA per gel. Purchased Criterion XT Bis-Tris 4-12 % gels (Bio-Rad) were run in a dedicated chamber (Bio-Rad) at 200 V with either XT-MOPS or XT-MES buffer (both Bio-Rad) depending on the desired mass resolution. To indicate mass separation the ROTI tricolor (Carl Roth) protein marker (10-245 kDa) was used.

Table 5: Formulation of 4x sample buffer according to [201].

10	mL	TOTAL	x4	
2.5	mL	1 M Tris-HCl pH 6.8	250	mM
0.5	mL	ddH20		
1	g	SDS	10	%
0.8	mL	0.1 % Bromophenol Blue	0.008	%
4	mL	100 % glycerol	40	%
2	mL	14.3 M β-mercaptoethanol	20	%

Table 6: Formulation of manually casted SDS-PAGE gels.

collection gel			separation gel	
4	%	Acrylamide	15	%
10	%	40 % acrylamide	37.5	%
12.5	%	1M Tris pH 8.8/6.6	37.5	%
75.4	%	ddH_20	22.9	%
1	%	10 % SDS	1	%
1	%	10 % APS	1	%
0.1	%	TEMED	0.1	%

3.4 FRET peptide assays

For SARS M_{pro} activity assays, peptide substrates were commercially purchased (Eurogentec), designed as SARS cleavage site analogues of twelve amino acids (P6-P6') with N- and C-terminal fluorophores to allow for Förster resonance energy transfer (FRET) (Table 7).

Initially, the freeze-dried peptides were dissolved in DMSO to 2 mM and stored at -20C°, protected from light. For sample preparation, the peptides were serially diluted in 50 mM HEPES, 10% (v/v) glycerol, 1mM DTT, pH 7.5 and final concentrations were adjusted by measuring A_{280} from five independent droplets, using ε 73,000 mol^{-1} cm^{-1}, the absorption coefficient of the HiLyte488 fluorophore as given by the manufacturer.

Assays were carried out on a 96-well plate reader (Infinite200, Tecan) with the following parameters: excitation wavelength 485 nm (bandwidth 9 nm), emission wavelength 535 nm (bandwidth 20 nm), gain 80, 10 flashes and 40 μs integration time. Further, Flat bottom Black polystyrol 96-well plates (Greiner) were used for fluorescence applications.

Table 7: FRET peptide substrates purchased to determine kinetic parameters of M_{pro}. Peptides were designed as cleavage site analogues with a FRET pair labelling, namely the fluorophor (F) HiLyte488 at the N-terminus and the Quencher (Q) QXL520 at the C-terminus.

SARS-CoV FRET substrate	Product
F-NSP4-5(N-term)	H-C(HiLyte™ 488)TS AVLQ-OH (7aa)
NSP4-5(C-term)-Q	H-SGFRK(QXL® 520)-NH2 (5aa)
F-NSP4-5-Q	H-C(HiLyte™ 488)TSAVLQSGFRK(QXL®520)-NH2 (12aa)
F-NSP7-8-Q	H-C(HiLyte™ 488)NRATLQAIASK(QXL®520)NH2 (12aa)
F-NSP8-9-Q	H-C(HiLyte™ 488)SAVKLQNNELK(QXL®520)-NH2 (12aa)
F-NSP9-10-Q	H-C(HiLyte™ 488)ATVRLQAGNAK(QXL®520)-NH2 (12aa)

Fluorescence coefficients k_{sx} and inner filter effect of the peptides were determined for each substrate according to previous protocols [202]. Briefly, to determine the specific substrate fluorescence the peptide was completely cleaved by incubation with SARS M_{pro} at a ratio of 1:1 for 20 h. Then, the peptide was nine times serially diluted (c_{sx} of 2 μM, 1 μM, 1.4 μM, 0.5 μM, 0.25 μM, 0.125 μM, 0.0625 μM, 0.0313 μM and 0.156 μM) and 100 μL were transferred into the wells.

The k_{sx} was determined as the slope of measured fluorescence in arbitrary fluorescence units (AFU) from 0.156 μM to 0.25 μM of peptide substrate according to (16).

$$\mathbf{k_{sx}} = \frac{\mathbf{AFU}}{\mathbf{c_{sx}}} \tag{16}$$

Furthermore, the inner filter effect was tested, according to a previously described method [202], to determine signal suppression of free fluorophore by the quencher of uncleaved peptides. The AFU of 98 μL uncleaved FRET peptide, S, at six substrate concentrations, c_x, (2 μM, 1.35 μM, 1 μM, 0.5 μM, 0.25 μM, 0.125 μM and 0 μM) was measured alone and with addition of 2 μL 0.25 μM free fluorophore F, namely F-NSP45 (N-term) (Table 7).

The inner filter correction factor for each concentration and peptide was determined with (17),

$$\textbf{inner filter corr.}\,(\textbf{at } \mathbf{c_x} \textbf{ of } \mathbf{S})\ [\%] = \frac{\mathbf{f(F + S_{cx})} - \mathbf{f(S_{cx})}}{\mathbf{f(F)}} \tag{17}$$

in which $f(F+S_{cx})$ is the AFU of the free fluorophore and the uncleaved FRET peptide at c_x, $f(F+S_{cx})$ is the AFU of the uncleaved FRET peptide at c_x alone and f(F) is the AFU of free fluorophore alone.

To determine protease to substrate reaction rates 20 µL SARS M_{pro} (final conc. c_{Mpro} of 0.5 µM) was mixed with 80 µL FRET peptide substrates of six different concentrations (final conc. $c_{X,S}$ of 2 µM, 1µM, 0.5 µM, 0.25 µM, 0.125 µM and 0 µM) and fluorescence in AFU was measured every 30 seconds for 5 minutes. To calculate kinetic parameters, triplicate measurements were performed.

The fluorescence coefficient k_{sx} (16) were used to convert the measured AFU into the concentration of cleaved peptide $c_{x,s}$ (18). The initial slopes of the progress curves at the beginning of the reaction corresponded to the initial rates k_{cat}/K_M (19). Finally, the turnover efficiency of the M_{pro} was obtained by dividing k_{cat}/K_M by enzyme concentration c_{Mpro} (20).

$$\frac{\mathbf{AFU}}{\mathbf{k_{sx}}} = \mathbf{c_{x,s}} \tag{18}$$

$$\mathbf{d\ c_{x,s}\ dt} = \frac{\mathbf{k_{cat}}}{\mathbf{K_M}} \tag{19}$$

$$\frac{\mathbf{k_{cat}}}{\mathbf{K_M}} * \frac{\mathbf{1}}{\mathbf{c_{Mpro}}} = \textbf{turnover rate } \mathbf{M_{pro}} \tag{20}$$

To test M_{pro} dimer-dependent activity, 20 µM FRET peptide substrate F-NSP4-5-Q (final concentration 2 µM) was incubated SARS M_{pro} of with 10 different concentrations (1.0 µM, 0.75 µM, 0.5 µM, 0.25 µM, 0.188 µM, 0.125 µM, 0.094 µM, 0.063 µM, 0.047 µM, 0.031 µM and 0 µM). The turnover rates were calculated as described above. The enzymatic activity k_{cat} and dimerization constant K_D were determined, assuming that the dimer of M_{pro} is the only active oligomer species, in a non-linear fitting (21) of turnover rates (Origin Pro) according to [64],

$$\mathbf{v_{max}} = \mathbf{k_{cat}[D]} = \mathbf{k_{cat}} \frac{\mathbf{K_D + 4C_{Mpro}} - \sqrt{\mathbf{K_D^2 + 8K_D C_{Mpro}}}}{\mathbf{8}} \tag{21}$$

in which v_{max} is the rate of enzymatic activity of M_{pro}, which is assumed to be similar to the turnover rate of the dimer $k_{cat}[D]$, and C_{Mpro} is the total enzyme concentration.

To determine the influence of salt conditions on enzyme activity, 20 µL M_{pro} (final concentration 0.1 µM) was incubated with 80 µM FRET peptide substrate F-NSP45-Q (final conc. 1 µM) in eleven concentrations of AmAc (10% glycerol, pH7.5 and 50 mM 165 mM 280 mM, 395 mM, 510 mM, 625 mM, 740 mM, 855 mM, 970 mM, 1085 mM and 1200 mM) or eleven concentrations of NaCl (50 mM HEPES, 10% Glycerol, pH7.5 and 0 mM, 120 mM, 240 mM, 360 mM, 480 mM, 600 mM, 720 mM, 840 mM, 960 mM, 1080mM and 1200 mM). Immediately after incubation, the fluorescence was measured every 60 seconds and the initial rates were determined from the initial slopes (AFU/s) as described above.

3.5 Native MS: Sample preparation

To prepare samples for native MS measurements, they were buffer exchanged into ESI compatible solution. Proteases M_{pro} were buffer exchanged into 250-500 mM AmAc, 1 mM pH 8 by two cycles of centrifugal gel filtration (Biospin mini columns, 6,000 MWCO, Biorad). The NSPs were buffer exchanged into 250-500 mM AmAc, 1 mM DTT, pH 8 by five rounds of dilution and concentration in centrifugal filter units (Amicon, 10,000 MWCO, Merck Millipore).

3.6 Native MS: nanoESI capillaries

Nano ESI capillaries were pulled in-house from borosilicate capillaries (1.2 mm outer diameter, 0.68 mm inner diameter, filament, World Precision Instruments) with a micropipette puller (P-1000, Sutter instruments) using a squared box filament (2.5 × 2.5 mm, Sutter Instruments) in a two-step program. Subsequently, capillaries were gold-coated using a sputter coater (Q150R, Quorum Technologies) with 40 mA, 200 s, tooling factor 2.3 and end bleed vacuum of 8×10^{-2} mbar argon.

3.7 Native MS: Obtaining mass spectra

Native MS was performed at an electrospray quadrupole time-of-flight (ESI-Q-TOF) instrument (Q-TOF2, Micromass/Waters, MS Vision) modified for higher masses [165]. Samples were ionized in positive ion mode with voltages applied at the capillary of 1300-1500 V and at the cone of 130-135 V. The pressure in the source region was kept at 10 mbar throughout all native MS experiments. For purpose of desolvation and dissociation, the pressure in the collision cell was adjusted to 1.3-1.5×10-2 mbar argon. Native like spectra were obtained at accelerating voltage of 10-30 V while for collision-induced dissociation these voltages were increased to 30-200 V. In ESI-MS overview spectra for NSP proteins, the quadrupole profile was 1-10,000 *m*/*z*. In tandem MS, for precursor selection, LMres and HMres were adjusted at 10-30 V collisional voltage until a single peak was recorded, then dissociation was induced.

To calibrate raw data, CsI (25 mg/ml) spectra were acquired and calibration was carried out with MassLynx (Waters) software. Data were analysed using MassLynx (Waters), Massign (by Nina Morgner [203]) and mMass (by Martin Strohalm [204]).

3.8 Native MS: NSP polyprotein processing

To start the processing reactions, NSPs and M_{pro} were incubated (ratio ~1:4–1:10). Three independent reactions were started in parallel and kept at 4°C until analysis. To acquire mass spectra at dedicated time-points, sample aliquots of 1-3 µL were withdrawn by means of a microliter syringe (5 µL, Hamilton) with flexible fused silica tubing (Optronis) and loaded into in house fabricated nanoESI capillaries, which were mounted on the nESI source, all within max. 2 minutes. Then, raw spectra were acquired in the first 300 scans (5 minutes).

To analyze the data the raw spectra were smoothed (2x5) in Masslynx 4.1 (Waters) and then NSP proteins were assigned to peak series. For each assigned mass species, SRs (using the relative intensities of peaks) were summarized and normalized. This was done independently for each spectrum. Finally, the average and standard deviation for SR and time-points was calculated from three independent spectra. Poorly resolved spectra were not included in the data analysis.

To determine mass fractions (MF), the SRs of assigned peaks for monomers and complexes were combined according to (22)-(27):

$$\mathbf{MF\ SARS\ NSP7 = SR\ [NSP7\ mon. + 2*NSP7\ dim. + NSP7{+}8(1{:}1) + 2*NSP7{+}8(2{:}2)]} \quad (22)$$

$$\mathbf{MF\ SARS\ NSP8 = SR\ [NSP8\ mon. + NSP7{+}8(1{:}1) + 2*NSP7{+}8(2{:}2)]} \quad (23)$$

$$\mathbf{MF\ SARS\ NSP9 = SR\ [NSP9\ mon. + 2xNSP9\ dim.]} \quad (24)$$

$$\mathbf{MF\ SARS\ NSP10 = SR\ [NSP10\ mon.]} \quad (25)$$

$$\mathbf{MF\ 229e\ NSP7 = SR\ [NSP7 + NSP7{+}NSP7{-}9]} \quad (26)$$

$$\mathbf{MF\ 229e\ NSP7{-}9 = SR\ [NSP7{-}9 + NSP7{+}NSP7{-}9]} \quad (27)$$

3.9 HDX mass spectrometry

Cleaved and uncleaved TGEV-CoV NSP7-8-NNE-His (50 pmol) were diluted 1:9 in 99% deuterated 20 mM Tris buffer (pH 8, 150 mM NaCl, 25°C) to start the exchange reaction. After three time points (1 min, 10 min and 60 min) the exchange reaction was quenched by 1:1 addition of ice-cold quench buffer (300 mM phosphate buffer, pH 2.3, 6 M urea), which decreased the pH to 2.3, and flash-freezing in liquid nitrogen. Each time-points represents a single measurement.

The samples were thawed and injected onto a cooled (0 °C) HPLC System (Agilent Infinity 1260, Agilent Technologies) equipped with a home packed pepsin column (IDEX guard column with an internal volume of 60 µL, Porozyme Immobilized Pepsin beads, Thermo Scientific) in a column oven (25 °C), a peptide trap column (OPTI-TRAP for peptides, Optimize Technologies) and a reversed-phase analytical column (PLRP-S for Biomolecules, Agilent Technologies).

Pepsin digestion was performed online at a flow rate of 75 µL per min (0.23 % formic acid in water) and peptides were trapped in the trap column. Peptides were eluted and separated on the analytical column using a 7 min gradient of 8-40 % solvent B (solvent A: 0.23 % formic acid in water, solvent B: 0.23 % formic acid in acetonitrile) at 150 µL per min. MS was performed using an Orbitrap Fusion Tribrid in positive ESI MS only mode (Orbitrap resolution 120K, 4 microscans).

Peptide identification was performed on non-deuterated samples using a elution gradient (27 min, 8-40 % solvent B) in data-dependent MS/MS acquisition mode (Orbitrap resolution 120K, 1 microscan, HCD 30 with dynamic exclusion of Top 20 N). Precursor and fragment ions were searched and matched against a local protein database in MaxQuant [205] with a minimum score of 20 for unmodified and 40 for modified peptides. DeutEx software (obtained from peterslab.org) was used to determine the deuterium uptake and create uptake plots. The peptide coverage map was plotted with MS Tools [206]. For analysis, cleaved and uncleaved NSP7-8-NNE-His were compared differentially regarding deuterium uptake according to (28).

$$\textbf{uptake difference} = \textbf{uptake}\,[\textbf{cleaved}] - \textbf{uptake}\,[\textbf{uncleaved}] \qquad (28)$$

Peptides spanning the cleavage site were ignored due to their absence in fully cleaved proteins. If uptake was lower in the polyprotein difference was positive. If uptake was lower in complex, the difference was negative. In absence of triplicate measurements only a preliminary data analysis could be performed and therein deuterium uptake was considered relevant when difference was at least ±0.5 Da and overlapping peptides showed the same trend [207].

3.10 MALDI-MS

For processing experiments, hCoV 229e NSP7-10 was mixed with SARS CoV M_{pro} in different ratios and for varying time. For complex formation, affinity chromatography and size exclusion chromatography purified SARS NSP7+8 and FIP NSP7+8 were cross-linked with 0.15 % Glutaraldehyde (Sigma-Aldrich) for 25 min. All samples were kept at 4°C until diluting them to 1 µM in MALDI matrix solution (Sinapinic acid 10 mg/mL in acetonitrile/water/TFA, 49.95/49.95/0.1, v/v/v). The samples were spotted (1 µL) on a stainless steel MALDI target plate. The MALDI-TOF/TOF mass spectrometer (ABI 4800, AB Sciex) equipped with a high-mass detector (HM2, CovalX) was used in linear TOF mode. For acquiring mass spectra (m/z 1,000 to 1,000,000 *m*/*z*) spots were ionized with a Nd:YAG laser (355 nm) and 500 shots per spectrum were accumulated. Obtained data were smoothed and analyzed using mMass (v5.5.0, by Martin Strohalm [204]).

4 Results and discussion

4.1 Quality control of non-structural proteins

To reach my set objectives, I needed to ensure a high quality of samples. Therefore, I carried out production and purification in-house, which gave me the possibility for a direct quality control and a short time-period between purification and analysis. In this section, I present the outcome of NSP sample production and quality control with native MS.

4.1.1 Coronavirus protease M_{pro}

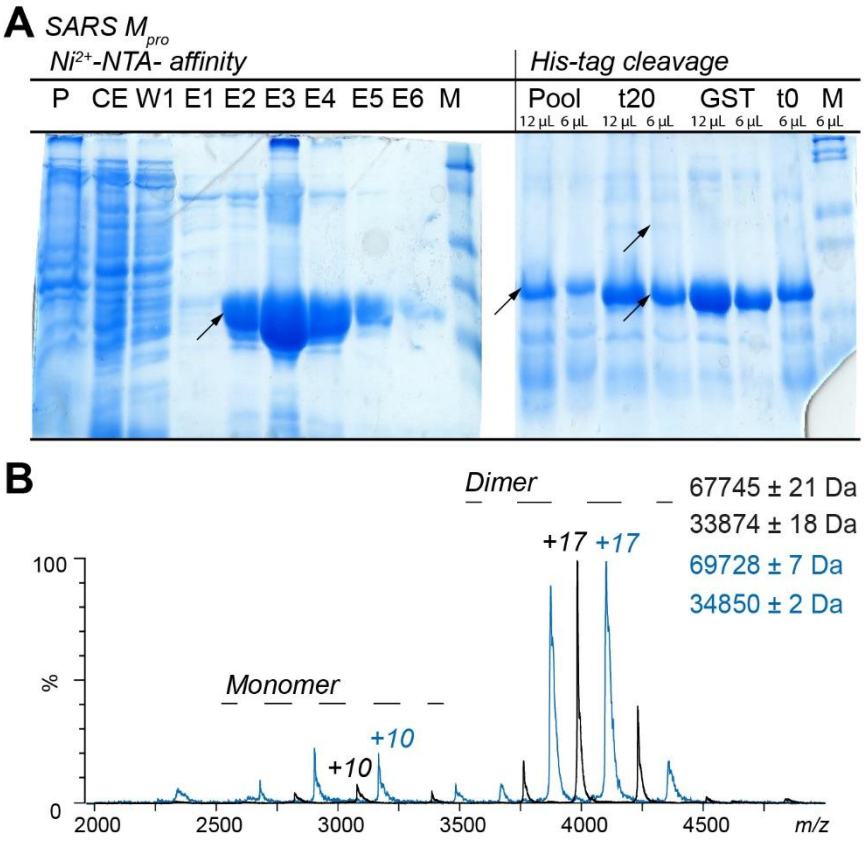


Figure 15: Purification and quality control of SARS M_{pro}. (A) SDS-PAGE of SARS M_{pro}-His purification with Ni^{2+}-affinity chromatography (left), cleavage of His-tag with PreScission and removal of PreScission with GST-affinity chromatography. Abbreviations: P, pellet; CE, crude extract; W1, first washing step; E1to E6, elution fractions 1 to 6; M, Spectra MS marker; Pool, pooled elution fractions; t20, 20 h cleavage of sample with PreScission; GST, sample after incubation with GST beads to remove PreScission; t0; cleavage of sample with PreScission. **(B)** Native mass spectra of 10 µm SARS M_{pro}-His (blue), as pooled fraction, and 10 µM SARS M_{pro} with authentic ends, as in GST fraction. Monomers and dimers indicated by independent charge envelopes. Main charge states labelled. Determined MW in Da given in legend.

The SARS coronavirus protease M_{pro} is the central protein for processing the SARS NSP polyprotein *in vivo* and *in vitro*. In order to obtain highly pure and enzymatically active M_{pro}, authentic N- and C-termini were required. Therefore a strategy was accessed and modified from Xue et al. [56], who had used an expression plasmid encoding SARS M_{pro} with a GST tag for solubility and a His-tag for affinity purification.

During expression, the N-terminal GST tag was cleaved at an original auto-cleavage site in the linker peptide creating an authentic N-terminus. After affinity purification, the His-tag was cleaved as well by addition of the homologous rhinovirus 3C protease (PreScission) with a slightly different substrate specificity creating an authentic C-terminus. Subsequently, the GST tagged Rhinovirus protease was removed by washing with GST beads. Finally, M_{pro} was flash frozen and stored at -80°C. To remove PreScission, the original protocol [56] included size exclusion and affinity chromatography steps that were omitted here, and instead the GST-bead wash was included. Nevertheless, M_{pro} reached sufficient purity as demonstrated in SDS-PAGE and native MS (Figure 15).

Several tests were run to check for protease quality. To control if activity was inhibited in AmAc electrospray solution, SARS M_{pro} was incubated with the specific substrate FPS4-5 in different concentrations of AmAc or NaCl (Figure S 2). The protease's activity was found to increase with increasing AmAc concentration but was inhibited with increasing NaCl. The inhibiting effect of NaCl on M_{pro} has been described before and is most likely based on interference of Na^+ or Cl^- with substrate attachment into the binding site. The results suggested that the bulkier molecules ammonium and acetate do not cause such an inhibiting effect.

In native MS, SARS M_{pro} was found consistently in a monomer-dimer equilibrium, which was concentration dependent. The concentration dependency of its activity has been described by other researchers and was confirmed here in a FRET assay (Figure S 1). The determined enzyme efficiencies at different concentrations showed clear dimer-dependent activity, as seen by the efficiencies disproportionally increasing with concentration. Furthermore, these acquired values allowed for fitting an equation with two variables, the K_D (K_D=0.012 ± 0.036 μM) of dimerization and k_{cat} (0.601±0.06 min^{-1}) for activity (Material and Methods: Section 3.4). These values deviate from published results [64, 208] obtained from similar assays, which here could not be reproduced. Throughout the data analysis, it was found that fitting requires more sampling points and less error than possible with this experimental setup at the used concentrations. Therefore, these experiments were not repeated.

During course of these investigations, also FIP and 229e proteases were checked with native MS and later used for *in vitro* processing experiments (Figure S 5). While specific activity was observed, the absolute efficiencies of FIP and 229e M_{pro} were not determined. However, native mass spectra indicated lower dimerization affinity and heterogeneity, probably induced by freeze damage.

4.1.2 NSP7-10 purification

Coronavirus polyproteins from the NSP7-10 region were produced as described here (Material and Methods: Section 3.2) and elsewhere before [83]. Hallmarks of this purification strategy were induction with anhydrotetracycline, expression at low temperatures, gentle sonication, and finally affinity and size exclusion chromatography.

Proteins of the NSP7-10 region were highly soluble upon cell lysis and highly pure after affinity chromatography (Figure 16). To counteract aggregation and disulfide bonding, proteins were strictly kept at low temperatures and in solutions with elevated ionic strength (250-500 mM NaCl) with freshly prepared reducing agent (2-5 mM DTT). The quality of the NSP7-10 was assessed mainly by SDS-PAGE and UV absorption at 280 nm. A sharp decline in quality was observed from five days on after expression. Therefore, experiments with NSP polyproteins were performed within five days after expression.

The design of NSP constructs was a crucial step because requirements of His-tag and the linker sequence varied according to the aim of the experiment. To create an authentic C-terminus by M^{pro} cleavage (///), it was sufficient to use the given linker in pASK35+ (resulting sequence: 'NSP'///SARGS-HHHHHH). However, to create an authentic N-terminus, additional amino acids were added (resulting sequence: HHHHHH-AVLQ///'NSP'). These additional amino acids represent the natural linker sequence of SARS M^{pro} auto-cleavage site NSP4-5.

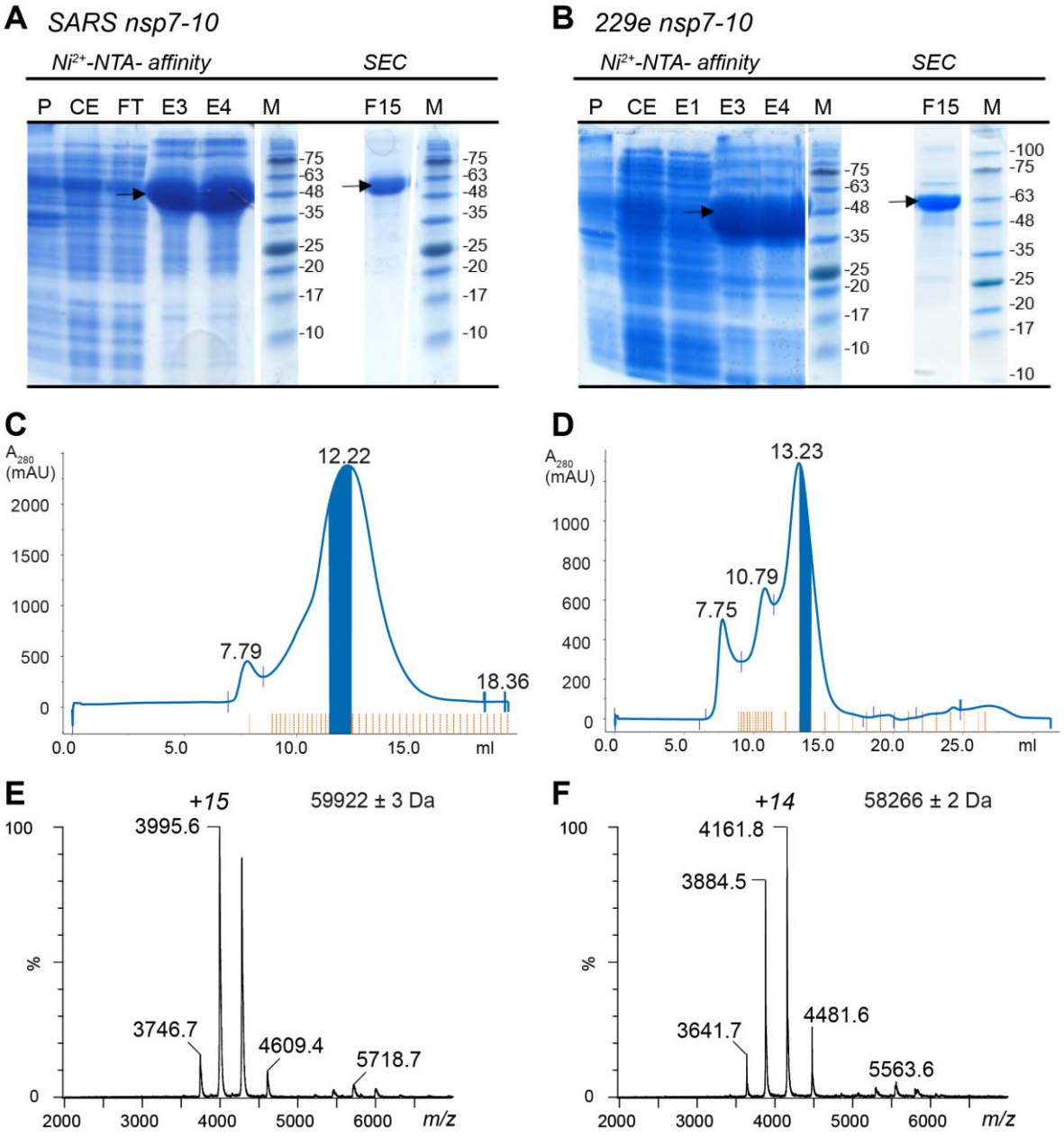


Figure 16: Purification of NSPs exemplified on 229e NSP7-10 and SARS NSP7-10. Polyproteins NSP7-10 from SARS **(A, C and E)** and 229e **(B, D and F)** produced in *E.coli*, purified by Ni^{2+}affinity chromatography, and SEC. **(A and B)** SDS-PAGE showing purity after chromatography steps. NSP7-10 protein bands indicated by black arrows. Abbreviations: P, pellet; CE, crude extract; FT, flow through; E3 and E4, elution fractions 3 and 4; M, Roti tricolor protein marker; F15, SEC elution fraction 15. **(C and D)** Size-exclusion chromatograms of NSP7-10 from SEC with Superdex200 (10/300) column showing A_{280} (blue line). Peaks labelled with elution volume (in mL). Elution fractions from main peak used for further analysis (blue bars). Peaks eluting earlier representing NSP oligomer species or co-purified *E.coli* proteins. **(E and F)** Native mass spectra of NSP7-10 showing purification to homogeneity and a mainly monomeric distribution at 15 µM. Peaks labelled with *m/z* and main charge states.

4.1.3 *Native MS of NSP10 revealed zinc binding*

Native MS revealed mass, charge and ionization efficiency, all indicative for sample quality. In the following paragraphs, I elaborate these factors and the obtained results regarding quality control of my sample proteins.

Analysis of the NSP proteins with ESI-MS allowed for an unmatched mass accuracy and therewith, an identification of the correct amino acid sequence. The determination of accurate molecular weight (MW) is preferably carried out in a denatured state. Often, denaturing the NSPs lead to strong aggregation and obtaining their mass spectra proved tedious. The mass determination from native solutions proved equally valid for NSP polyproteins as well as for their processing products. The MWs were unequivocally determined corresponding to their theoretical average mass, calculated from amino acid sequences. In the appendix of this thesis, listed are used NSP sample proteins, their theoretical ($MW_{theo.}$) and experimental MW ($MW_{exp.}$) (Table S 1: Molecular weights (MWs) of NSPs, their cleavage products).

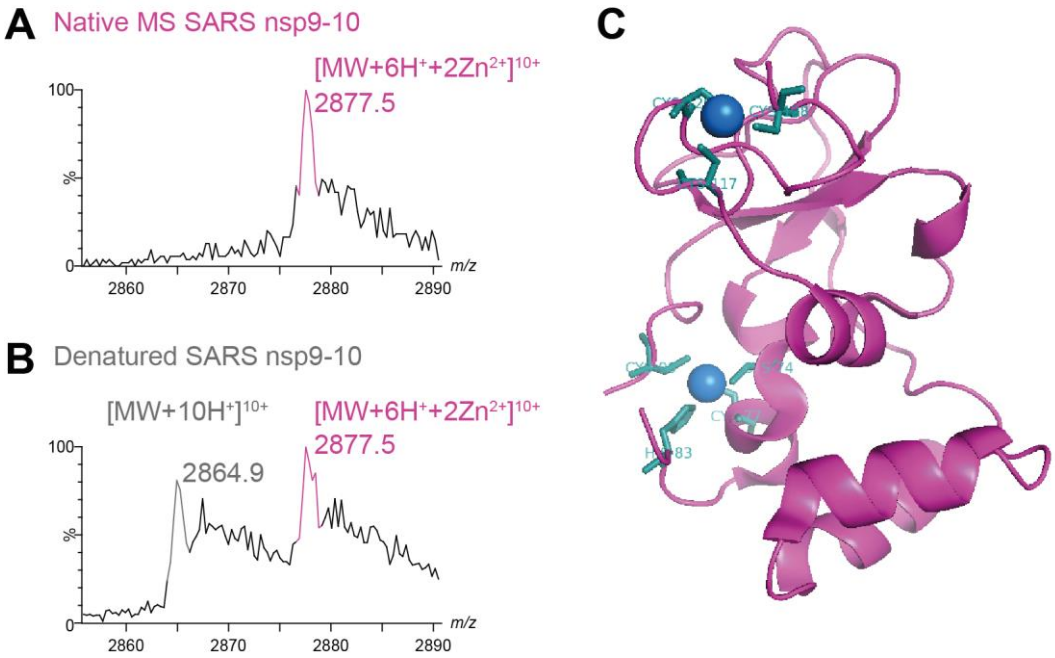


Figure 17: SARS NSP10 zinc binding revealed by ESI-MS. **(A)** Native mass spectrum of SARS NSP9-10 showing a clipping of the molecular ion +10 charge state. Highlighted (pink) is Zinc bound NSP10, deviating +130 Da from the amino acid sequence of NSP9-10. **(B)** ESI mass spectra of denatured SARS NSP9-10 showing partial loss of zinc upon denaturation. Additionally to the zinc bound ion signal (pink) an additional signal (grey) corresponding to the theoretical ion mass without two zinc. **(C)** SARS NSP10 structure from X-ray crystallography (PDB 2G9T, [94]) showing binding of two Zn^{2+} molecules. Visualized with PyMOL. Highlighted are NSP10 (pink), bound Zn^{2+} ions (deep-blue), and the coordinating amino acid side chains (teal).

Exceptions of the theoretical mass were found for SARS and 229e NSP10, and for all polyproteins containing the NSP10 domain. They were consistently determined with a mass shift of 130 Da. This mass shift was already suggested by the presence of two zinc finger

domains within NSP10, each coordinating one zinc molecule with the mass of 65 Da [94]. To prove this experimentally, proteins containing the NSP10 domain were carefully denatured in solution to trigger losing their zinc upon structural disintegration. In fact, the mass of the denatured protein was found reduced by 130 Da, corresponding to the loss of two Zn^{2+} (Figure 17).

Exemplified below is the observed loss of zinc from the sample protein NSP9-10-QA-His as analyzed by ESI-MS. This protein has an average theoretical mass of MW=28639.6 Da and forms the molecular ions (29) and (30), respectively.

$$[\mathbf{M + 6H^{+} + 2Zn^{2+}}]^{\mathbf{10+}} = \frac{\mathbf{28639.6\ Da + \ 6\ Da + 130\ Da}}{\mathbf{10\ z}} = \boldsymbol{theoretical}\ \mathbf{2877.6}\ \boldsymbol{m/z} \quad (29)$$

$$[\mathbf{M + 10H^{+}}]^{\mathbf{10+}} = \frac{\mathbf{28639.6\ Da + \ 10\ Da}}{\mathbf{10\ z}} = \boldsymbol{theoretical}\ \mathbf{2865.0}\ \boldsymbol{m/z} \quad (30)$$

The theoretical ions corresponded well to the actual signals detected from buffered solution, 2877.5 *m*/*z*, and from denaturing solution, 2864.9 *m*/*z*. These signals were evidence that the additional mass detected, in fact, derived from binding of two zinc to NSP10.

Remarkably, without denaturation, NSP10 was exclusively present in a Zinc bound state, even though there was no additional Zn^{2+} present in expression medium or purification buffer. Detection of zinc in NSP10 was further evidence for the correctly folded protein in solution as well as the conservation of structural elements in the gas phase.

4.1.4 *Mass-to-charge states of NSPs*

Electrospray based charge states of proteins depend on their collisional cross section in the gas phase. To test for extremely high or low charge states, the average charge states of NSPs were compared in a typical mass to charge plot (Figure 18). The charging of NSPs was slightly below the Rayleigh limit for spherical proteins, z_R [209]. Therefore, eminent protein unfolding was ruled out, because this would have resulted in an adoption of much higher charge states.

In fact, most charge states were within 65 to 100%of z_R, which, as described elsewhere, is consistent with the charge residue mechanism in ESI ionization [154]. Nevertheless, some of the smaller NSP proteins had charge states even below the 65%, which shows that for small globular proteins, these approximations are not as valid.

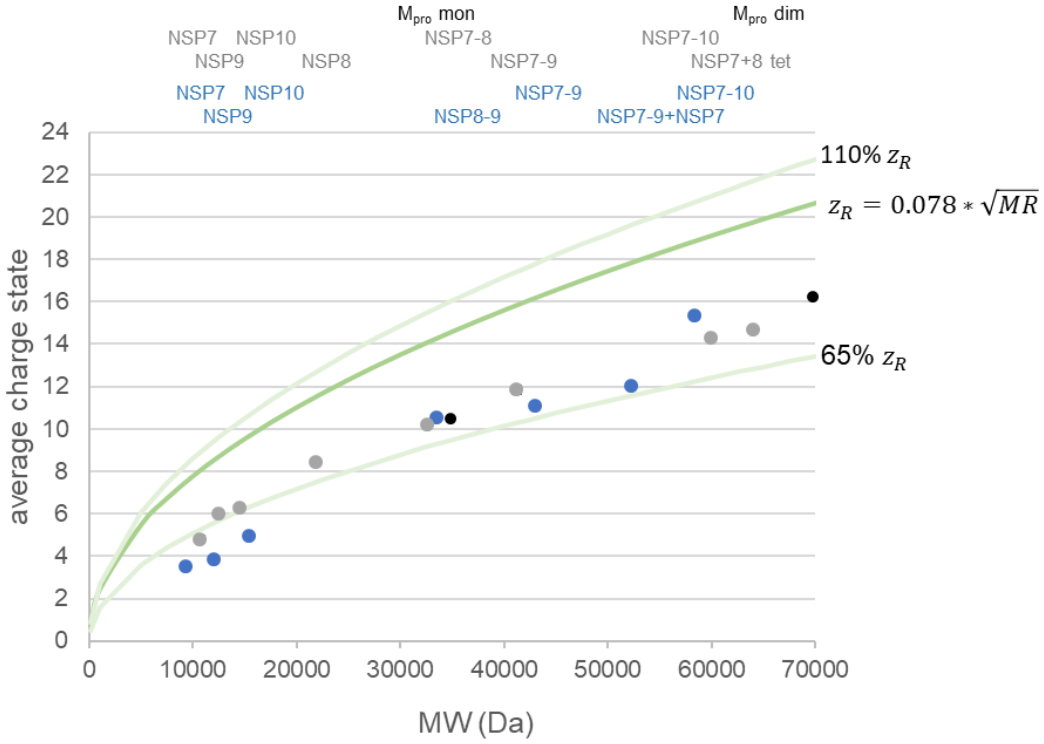


Figure 18: Mass-to-charge plot of coronavirus NSPs. Average charge states plotted against MW. All average charge states were in good approximation below the theoretical Rayleigh limit (green line) for spherical proteins, which indicates the maximum charge z_R that can be adopted. NSP proteins of 229e (blue), SARS (grey) and SARS M$_{pro}$ (black) analyzed by native MS in 250-500 mM AmAc, pH 8. Sample proteins indicated above. Standard deviation for the average charge (SD 0.02 to 3.97 %) is depicted, but too low to visibly appear in graph.

4.1.5 Signal response ratio of NSP7 to NSP10

If the concentration of different analytes in a mixture is equal, the difference in signal response is only dependent on the response factor. Since concentrations of the single NSPs from cleaved polyproteins were equimolar, signal ratio of monomers provided a good estimation for deviating response factors.

In this work, the ratio between NSPs was determined by two different approaches, exemplified here for completely cleaved SARS NSP7-10 (Figure 19). First, signal responses were directly obtained from the peak intensities of the monomers of NSP7, NSP8, NSP9 and NSP10. However, the relative peak intensities (rel. int.) between them were largely different. This could be based on response factors, but also that a certain fraction of the NSPs formed complexes. Therefore, they should not be directly compared but observed in a semi-quantitative fashion, where increasing and decreasing ratios indicate a relative change in concentration.

Second, ratios of mass fractions (MFs) were obtained by using the relative peak intensities of the monomers and add the relative intensities of all non-covalent complexes, which contain the respective monomeric NSP subunits (Equations listed in 3.8). The mass fractions of NSP7, NSP8, NSP9 and NSP10 were found evenly distributed, indicating that they nicely represent the equimolar presence of NSPs, if monomeric or bound in a complex. This result shows that a plot of the mass fractions could be used to observe relative concentrations directly from the ratios.

Furthermore, instrument parameters and response factors appear to have less influence than expected. However, previously it was shown that with a similar instrument, larger species give more intense signals. Here, the tested NSPs (~9.5-15 kDa) can have different ion efficiencies than larger species, e.g. NSP7-10 and M_{pro} (~59 kDa and ~69 kDa), and should therefore not be compared directly.

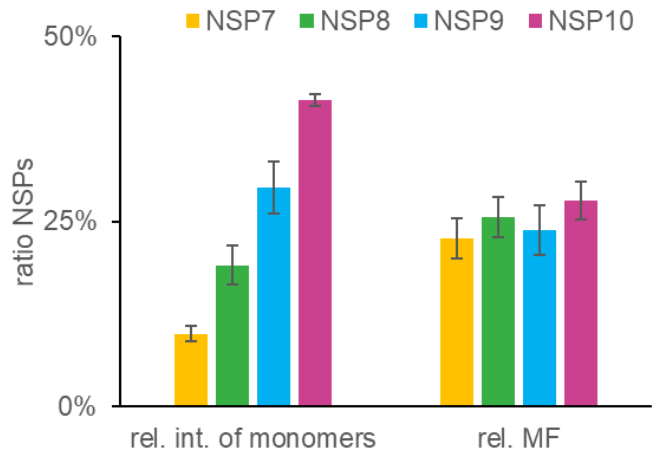


Figure 19: Signal response ratio of SARS NSP7-10 domains. Bar diagrams of rel.in. ratios of NSP7, NSP8, NSP9 and NSP10. The rel. int. ratio was determined by summing up SR directly assigned to the NSP monomers (AVG.±SD, N=3; 10±1%, 19±3%, 30±4% and 41±1%). The ratio of mass fraction (MF) is the SR ratio, corrected for all non-covalent complexes (AVG.±SD, N=3; 23±3%, 26±3%, 24±3% and 28±3%).

4.1.6 Expression of NSP6

The membrane protein NSP6 was of interest because of its clamped position within the polyprotein, between M_{pro} and NSP7-10. A fully folded recombinant NSP6 could add value my investigations. Here I briefly present work aimed to produce recombinant NSP6 for *in vitro* analysis (Figure 20).

Production and purification of full-length membrane proteins is still an ongoing challenge for biochemists, but shortened protein versions can simplify experiments. Studying literature and performing *in silico* analysis of the amino acid sequence suggested that NSP6 comprises of seven membrane domains (aa1-aa241) followed by an approximately 6 kDa folded endo-domain NSP6c (aa242-aa292). With this knowledge at hand, two protein constructs of NSP6c were cloned, both N-terminally tagged for solubility and purification, with His-NSP6c and His-GST-NSP6c, respectively. Both constructs were transformed in *E.coli* and while expression of His-NSP6c was either not successful or the product too small for detection in SDS-PAGE, expression of His-GST-NSP6 possibly gave a positive result, as concluded by the low resolution in SDS-gel. However, subsequent attempts to solubilize the complete His-GST-NSP6c protein were not successful.

Eventually, purification lead to a truncated protein. Native MS of the elution fraction assigned the truncation product to aa1-aa230, encompassing GST tag and parts of the linker. The assignment was based on the mass species detected being equal to the theoretical mass of aa1-230 (theoretical MW 26.97 kDa; measured MW 26.95 kDa). Additionally, a monomer-dimer mass distribution (26.95 kDa; 53.93 kDa) was observed, indicative of GST, which is also known to dimerize. The question was raised, if aa1-230 arose in fact from truncation or from expression of a vector lacking the NSP6c sequence. Analyzing the vector sequence helped to exclude latter, because an empty vector would give rise to a much longer product (aa1-355aa). Therefore, it is likely that the protein was completely expressed, due to the detection of the putative full-length His-GST-NSP6c in the pellet fraction in SDS-PAGE. Ultimately, the results suggested that production of NSP6c requires optimization for example by including other NSP6 endo-domains in the construct.

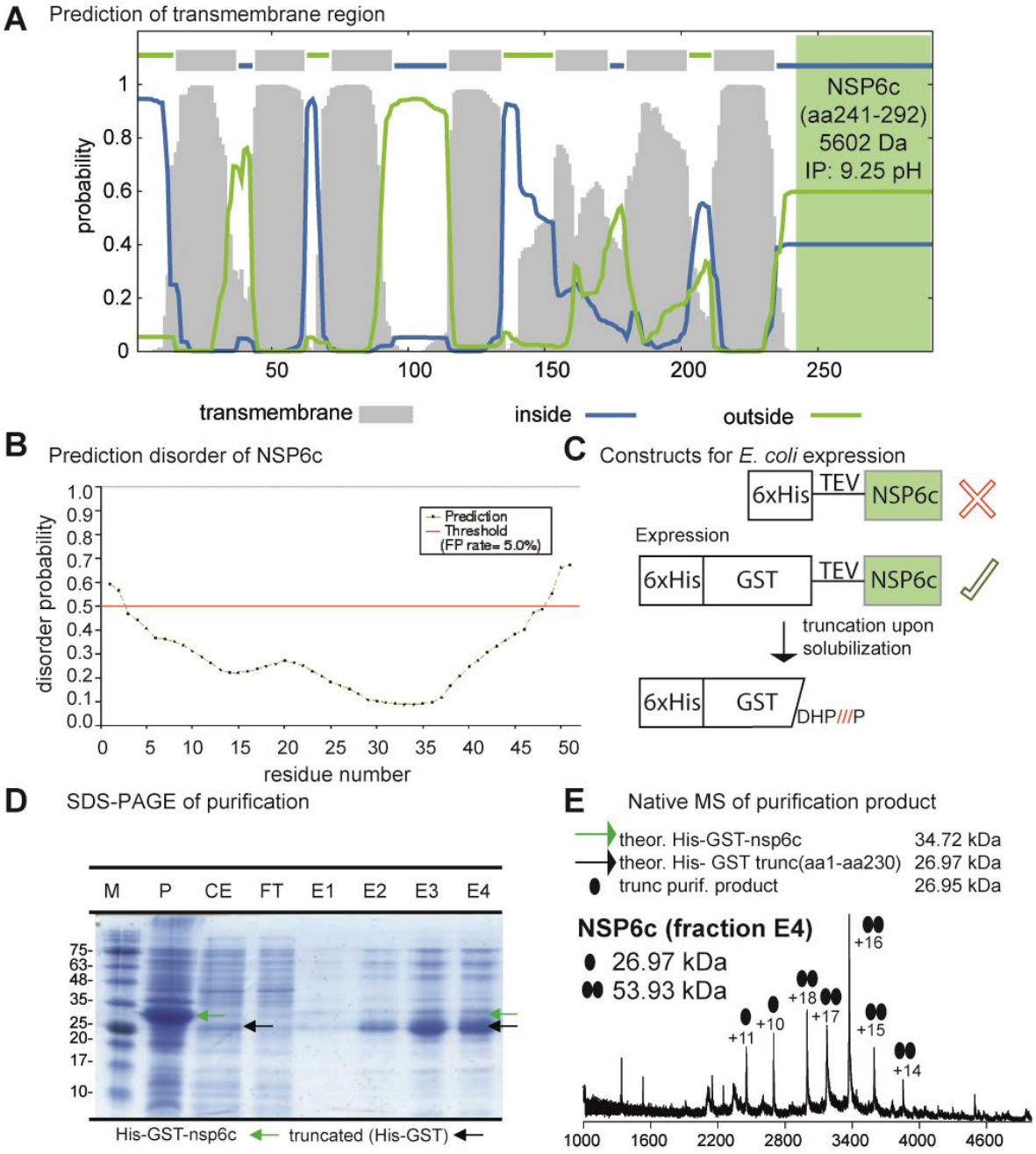


Figure 20: NSP6c is an insoluble endo-domain.(A) *In silico* analysis of NSP6 with transmembrane domain predictor (TMHMM server, v2.0 , [210]). NSP6c(green box) is a putative 51 amino acid long C-terminal endo-domain of NSP6. **(B)** *In silico* disorder prediction of NSP6c (PrDos, [211]). Black line representing amino acid chain and red line limit for disorder. Central amino acid sequence of NSP6c predicted to be of ordered structure. **(C)** Schematic of the gene constructs His-NSP6 and His-GST-NSP6. Putative truncation product of His-GST-NSP6 indicated below. **(D)** SDS-PAGE (15%) of His-GST-NSP6 purification steps. Possibly complete protein (green arrow) in the pellet and truncated version (black arrow) in the soluble fraction. Abbreviations: M, Roti tricolor protein marker; P, pellet; CE, crude extract; FT, flow through; E1 to E4, elution fractions 1 to 4; **(E)** Native MS of buffer exchanged Ni^{2+} affinity purification fraction E4 shows monomer-dimer distribution of a 26.97 kDa and 53.93 kDa mass species. This mass species equals theoretical MW of aa1-aa230 of His-GST-NSP6 (black arrow), encompassing the GST molecule.

4.2 Processing of CoV NSP7-10 region

In order to replicate, SARS-CoV requires processing of its polyprotein by M_{pro} to release NSP domains, which form the replication-transcription complex. The order of release is dependent on the proteolytic activity of M_{pro} at cleavage sites located at NSP inter-domain junctions. One objective of this thesis was to learn about the order of release from polyprotein NSP7-10 of SARS and 229e. In a first set of experiments, I set up *in vitro* processing with different substrates, starting with cleavage-site peptide substrate analogs in a fluorescence assay and continuing with full-length NSP proteins in Native MS. In this section, I present results about the processing of FIP NSP7-9 region and SARS as well as 229e NSP7-10 region.

4.2.1 Processing of cleavage site analog FRET peptides

Oftentimes, peptide assays had been used to test the activity of M_{pro}. In order to relate to the results of these investigations, FRET peptide substrates (FPS) were designed representing the cleavage sites at SARS polyprotein inter-domain junctions of NSP4-5, NSP7-8, NSP8-9 and NSP9-10. The FPS encompassed 12 amino acids of the P6'-P6 cleavage sites labelled with a fluorophore and a quencher and were termed accordingly FPS4-5, FPS7-8, FPS8-9 and FPS9-10. Upon M_{pro} cleavage, quencher and fluorophore dislocated leading to proportionally increasing fluorescence. The proteolytic activity for each substrate was determined as a function of increasing fluorescence (Figure 21).

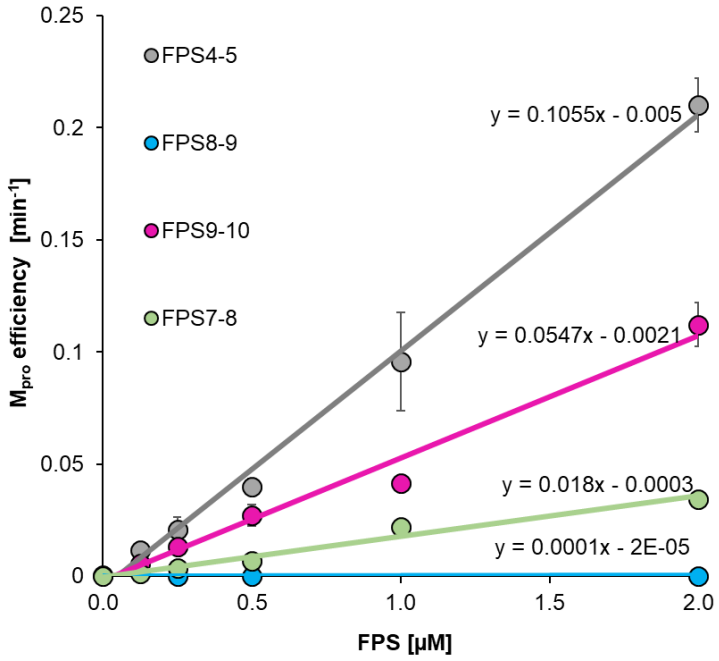


Figure 21: M_{pro} assay with FRET peptide substrates. Apparent (k_{cat}/K_M) efficiency of M_{pro} plotted against FRET peptide substrate (FPS) concentration. Specific turnover number derived from the slopes of apparent efficiency. Error bars indicate for standard deviation from triplicate measurements. The FPS specific fluorescence coefficients and influence of inner filter effect were tested as well (Figure S 3).

Table 8: Substrates and their relative efficiency for M_{pro}. Given is the name of the substrate, the amino acid sequence P6'-P6 at NSP inter-domain junction cleavage site, relative efficiency as determined in FRET peptide protease assay. Highlighted (blue) is a conserved NNE motif of the non-canonical NSP8-9 cleavage site.

Name	Amino acids	Rel. efficiency (%)	Efficiency k_{cat}/K_M (μM^{-1} min^{-1})
FPS4-5	TSAVLQ///SGFRK	100	0.105±0.005
FPS7-8	NRATLQ///AIASK	17.1	0.018±0.003
FPS8-9	SAVKLQ///NNELK	0.1	0.0001±2x10^{-5}
FPS9-10	ATVRLQ//AGNAK	51.8	0.0547±0.002

The substrate FPS4-5 is analogue to the N-terminal auto-cleavage site of SARS M_{pro}. Between the substrates tested here, FPS4-5 was most efficiently cleaved with a turnover rate (k_{cat}) of 0.105±0.005 $^{-min}$. This turnover rate of FPS4-5 is in line with recently published results, deriving from a similar method [208].

Compared to FPS4-5 (100 %), substrates FPS7-8 (17.1 %) and FPS9-10 (51.8 %) were less efficiently cleaved while FPS8-9 (0.1 %) was virtually non-cleavable. Previously, other researchers had also found considerably low activity at this site. The NSP8-9 cleavage motif has a *Coronaviridae*-wide conserved NNE motif at P'1-P'3. This gave rise to the hypothesis that a putative low activity had a functional role by keeping NSP8-9 domains intact. To see if these results from the peptide assay hold true for more native-like substrates, full-length NSP7-10 proteins were tested.

4.2.2 Processing of full-length SARS NSP7-10

In order to investigate polyprotein processing, peptides containing only a short stretch of amino acids had been analyzed. In the peptide assays, the role of other structural features beyond the sequence had been left out. To evaluate processing with a more native-like substrate, full-length NSP proteins were cleaved *in vitro* and cleavage products were observed with native MS and SDS-PAGE.

Initially, processing of SARS NSP7-10 was investigated with native MS. This protein contained an N-terminal His-tag, which mostly prevents a known interaction of NSP7 with NSP8. This tag was beneficial, because for processing only covalent product were relevant and therefore, it simplified data analysis.

To establish the native MS approach, a range of M_{pro} to polyprotein ratio (1:5-1:8) was determined that allowed on the one hand, complete processing within 20 h and on the other hand, prevented masking of product peaks by M_{pro}. Furthermore, the absolute concentration was limited to 20 µM to reduce artifacts of molecular crowding. The reactions were carried out at 4 °C to keep the protein products stable for 20 h.

Upon native MS of cleaved NSP7-10, peaks in the mass spectra were assigned to various protein species (Figure 22). They sort into three categories: substrate and protease, intermediate products representing partially cleaved substrate and final products encompassing NSP monomers and protein complexes.

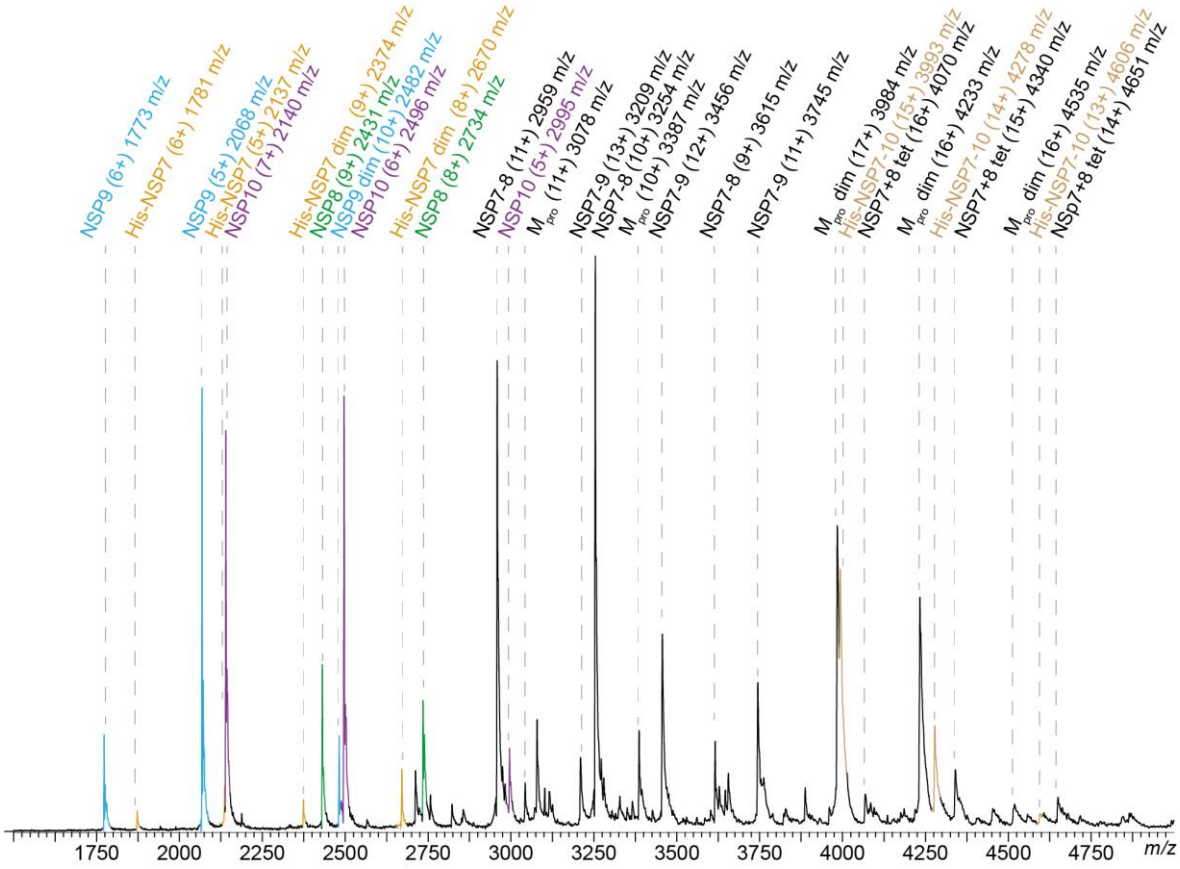


Figure 22: Proteins assigned to peaks in native MS of SARS His-NSP7-10 cleavage. Native mass spectrum of SARS His-NSP7-10 polyprotein cleaved by SARS Mpro for 6 h, when cleavage is not yet complete. The protein species can be sorted into three categories, substrate and protease (Mpro and NSP7-10), intermediate products (NSP7-8 and NSP7-9) and final products encompassing NSP monomers and protein complexes (NSP7, NSP7 dimer, NSP8, NSP7+8 tetramer complex, NSP9, NSP9 dimer and NSP10). In order to analyze the time-course of processing, protein species had to be assigned to peaks in the mass spectra. The NSP7 dimer peak (2374 m/z) has been unintentionally left out of processing analysis. However, this peak had a maximum of 1.5% of total signal intensity and therefore, was not significant for the results. Some peaks were not assigned because they were masked at other time-points and therefore, not suitable for analysis.

When acquiring mass spectra at different times during processing, signals for the three categories dominate successively (Figure 23); in the beginning (~1h) polyprotein and M_{pro}, during processing (~6 h) intermediate products and at the end-point (~20 h) final products. Covalent and non-covalent products were distinguished with CID-MS (Figure 26).

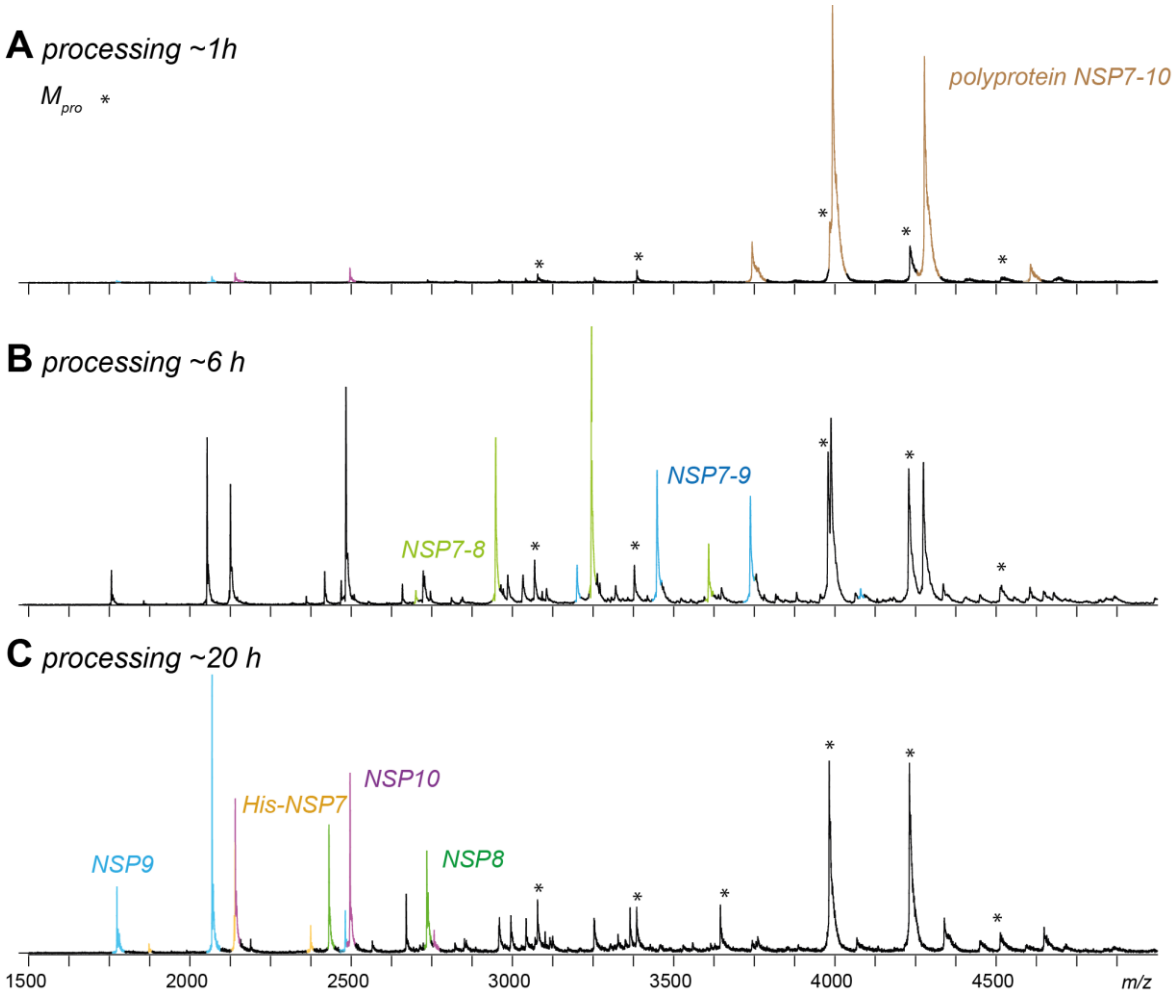


Figure 23: SARS NSP7-10 processing in native MS: Exemplary spectra. Exemplary native mass spectra illustrate processing from polyprotein to NSP end-products. For *in vitro* processing, 1.25 µM SARS-CoV M_{pro} and 12.5 µM SARS NSP7-10-His (~1:10 ratio) were incubated at 4°C in 250 mM AmAc, 1 mM DTT at pH 8. After mixing the components, samples were injected into an electrospray capillary and native MS spectra were recorded at indicated time points. **(A)** Mass spectrum at 1 h showing NSP7-10 (beige). Additionally indicated are peak signals of M_{pro} monomer and dimer (asterisk), as well as low intensity signals of NSP9 (blue) and NSP10 (pink). **(B)** Mass spectrum at 6 h with dominant peaks assigned to intermediate products NSP7-8 and NSP7-9. **(C)** Mass spectrum at 20 h showing dominant peaks assigned to final cleavage products His-NSP7 (yellow), NSP8 (green), NSP9 and NSP10. (**Table S 1: Molecular weights (MWs) of NSPs, their cleavage products and complexes.**)

The order of several reactions became evident from monitoring peak intensities over time. Following the time-course of protease and substrate, the relative signal intensity of M_{pro} remained stable while NSP7-10 steeply decreased (Figure 24), depicting that M_{pro} depletes NSP7-10 and thereby converts it in cleavage products. After 20 h, no signal of NSP7-10 was detected anymore and therefore, cleavage was considered complete.

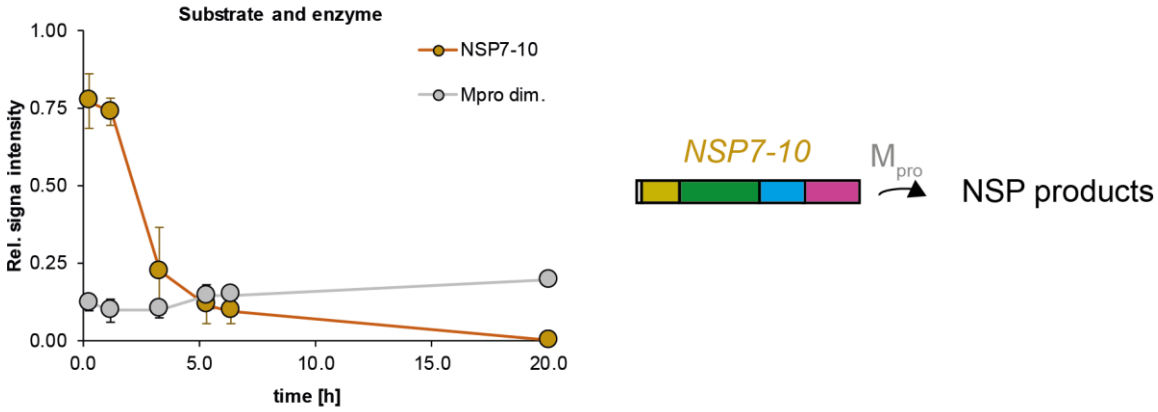


Figure 24: SARS NSP7-10 processing in native MS: Signal over time of protease and substrate. Relative signal intensity over time shows that NSP7-10 depletes and M_{pro} remains stable. Relative signal intensities obtained in one experimental session and plotted in different graphs for the purpose of clear representation (Other graphs from this analysis: Figure 25). All data obtained in one experimental session but shown in different graphs for clear presentation. Symbols and arrows illustrate conversion of NSP mass species as concluded from the graphs. Error bars show Std. Dev. (N=3) for relative intensity. Time points (AVG±SD, N=3): 0.25±0.1 h, 1.2±0.1 h, 3.3±0.2 h, 5.3±0.3 h, 6.4±0.3 h±0.5 h.To obtain the order of NSP7-10 cleavage, the ratio of the NSP products was observed. As shown for the signal response (SR) ratio (Chapter 4.1.5), a change in relative intensity shows a change of relative concentration and the ratio of mass fractions shows the ratio of concentrations.

Observing the relative signal intensities of NSP monomers, the highest increase was shown by NSP10 followed by NSP9, while only a slow increase was shown by NSP7 and NSP8 (Figure 25 A). The relative increase suggested an order of monomeric NSP products present in solution, which was difficult to correlate to NSP cleavage from the polyprotein. Notably, at the end-point of processing (~20 h) the relative intensity of NSP monomers remained largely different, despite being equimolar present (Figure 19).

Non-covalent interactions of NSPs were identified, which resulted in a bias in the relative monomer intensities. Correcting for the non-covalent interactions, the mass fractions (MFs) were calculated and found evenly distributed at the end-point of processing (~20 h), when cleavage was considered complete (Section 4.1.5). Therefore, ratio of MFs at earlier time points should reveal the relative NSP concentration. In fact, there was clear order of MFs during processing (Figure 25 B). The highest MF increase was shown by NSP10 followed by NSP9 and then equally by NSP7 and NSP8. These results revealed the order of NSP cleavage off of the

polyprotein and therefore, order of M_{pro} cleaving first the NSP9-10 site then the NSP8-9 site and last the NSP7-8 site.

To further specify, at 3.3 h the relative MFs between NSP10, NSP9, NSP8 and NSP7 were 100%, 66%, 36 % and 36 %, respectively. If NSP10 is normalized to a relative efficiency as determined in the peptide assay for FPS9-10 (52 %) and a strict cleavage order from the polyprotein is assumed, then the relative cleavage efficiencies in the full-length NSP7-10 between NSP9-10, NSP8-9, NSP7-8 were 52%, 34% and 16-17 %. In fact, for NSP7-8 this approximation well agrees with the result from the peptide assay for FPS7-8 (17.1 %). However, many factors could influence readouts of efficiencies in such a multi-dynamic system and protein constructs with a single cleavage site would be better suited to determine such an exact number.

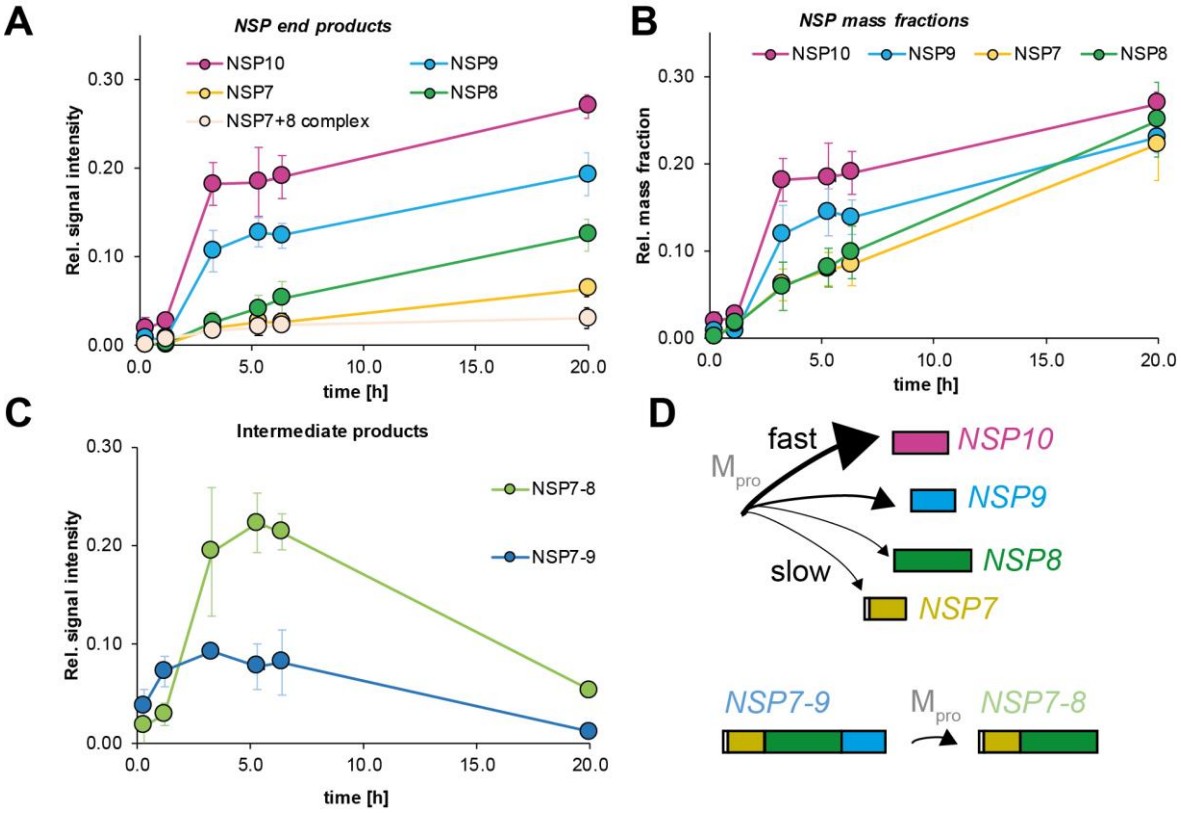


Figure 25: SARS NSP7-10 processing in native MS: Signal over time. Time course of SARS His-NSP7-10 *in vitro* processing (Figure 23). Shown in the graphs are relative signal intensities at different time-points. All data obtained in one experimental session and plotted in different graphs for the purpose of clear representation. **(A)** Intensity of NSP monomer subunits increase when cleaved from polyprotein. The protein complex from NSP7+8 increases as NSP7+8 is increasingly detected. **(B)** Intensities of NSPs, corrected for non-covalent complexes containing NSPs, in a plot of mass fractions over time clearly shows the order of NSP appearance. **(D)** Intermediate products NSP7-8 and NSP7-9 appearing sequentially. NSP7-9 is quickly superseded by NSP7-8, which becomes the most dominant species overall. These subsequent reactions shows that processing converts NSP7-10 first to NSP7-9 and then to NSP7-8, before at last NSP7-8 site is cleaved into NSP7 and NSP8. Error bars depict standard deviation (N=3). Time points (AVG±SD, N=3): 0.25±0.1 h, 1.2±0.1 h, 3.3±0.2 h, 5.3±0.3 h, 6.4±0.3 h±0.5 h.

Additionally, intermediates NSP7-9 and NSP7-8 were detected at high intensities, indicating that NSP7-10 had been converted into these mass species (Figure 24 C). These intermediates must have been generated by cleavage of NSP9 and NSP10, which is consistent with results from the MFs of the NSPs. Strikingly, relative signal intensities show that at first NSP7-9 was the main intermediate but then during the course of processing was replaced by NSP7-8, which became the most intense and longest lasting signal at 6 h. This time course of intensities clearly demonstrates that NSP7-10 is predominantly cleaved to NSP7-9 and only then to NSP7-8 (Figure 25 D).

In parallel to processing, tandem MS was performed to avoid misinterpretation due to non-covalent and covalent products (Figure 26). Particularly, NSP7-8 intermediate and NSP7+8(1:1) dimer would have signaled at similar *m*/*z*.

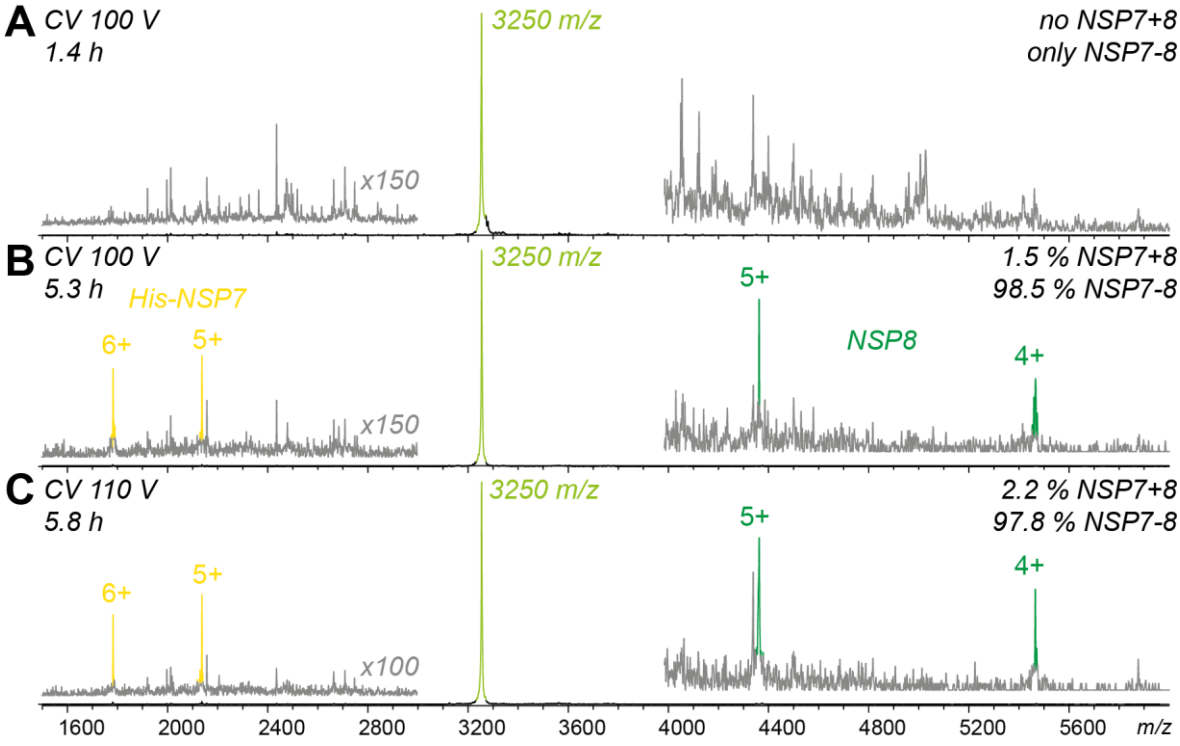


Figure 26: SARS NSP7-10 processing in native MS: CID of NSP interactions. Product ion spectra (black) of His-NSP7-8/NSP His-7+8 precursor (3250 *m*/*z*, 10+, light green) at three time-points of NSP7-10 processing (ESI-MS spectra of processing: Figure 23). Overlaid product ion spectra (grey) magnified at indicated ratio. Dissociation triggered in Q-TOF tandem MS with CV of 100V at 14 µbar argon. The ratio of His-NSP7-8 to NSP His-7+8 (1:1) is given, based on the ratio of intensity of precursor ion peak to NSP7 product ion peaks for **(A)** 1.4 h ,**(B)** 5.3 h and **(C)** 5.8 h. After 5.8 h, when the precursor ion is the dominant peak in the ESI-TOF spectrum, only 2.2 % of the molecular ion is dimeric, and 97.8 % is covalently bound.

CID spectra at three different time points of processing shows that the main molecular ion (10+, 3250 *m*/*z*) in question contains almost exclusively (97.2%) NSP7-8 intermediate. Additionally CID was used to unequivocally identify three charge states of a NSP7+8 (2:2) complex (Figure S 6, See section 4.3.1 for detailed analysis of NSP7+8 complexes). The dimer and tetramer

complexes were found at low intensity, which is a result of N-terminal His-tag on the NSP7 domain, sterically preventing a higher binding affinity. Nevertheless, the fraction of bound NSP7+8 was sufficient to introduce bias to the monomer ratio between NSPs, which explains why signal intensities of NSP7 and NSP8 never reached the ones of NSP9 and NSP10. In the end, the CID was important to correct for this bias by identifying non-covalent interactions, which were also used to calculate mass fractions.

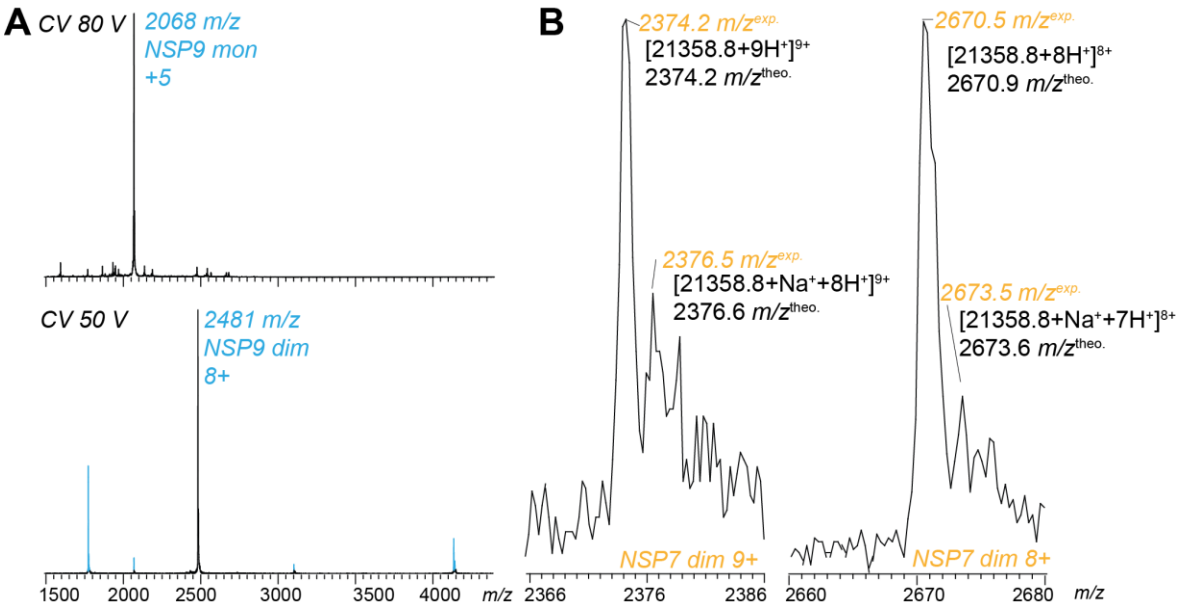


Figure 27: Homo-dimers of NSP7 and NSP8 identified in SARS His-NSP7-10 processing. **(A)** Collision induced dissociation to identify ions of monomeric and dimeric mass species of NSP9. Product ion spectra of 2068 *m/z* (top) precursor shows onset of backbone fragmentation at 80 V CV, while 2481 *m/z* precursor (bottom) shows dissociation at 50 V CV. Q-TOF tandem MS at 14 µbar argon. **(B)** Zoom of two charge states of a peak series assigned to NSP7 dimer in ESI-MS spectra of processing. Sodium adducts of experimental m/z^{exp} are well in line with theoretical m/z^{theo} that are expected from a NSP7 dimer. A monomeric NSP7 could not have adducts at the indicated m/z^{exp}. For affiliating native MS spectra see Figure 22.

Furthermore, two homo-dimer mass species were identified in MS analysis. The dimer of NSP9 had been repeatedly reported, but here, initially, only one peak series was assigned to NSP9 monomer. CID was performed to identify the molecular ions of the assigned peaks and while the main ion (2068 *m/z*, 4+) did not show any dissociation, in fact, the highest m/z ion (2481 *m/z*, +8) showed dissociation into two NSP9 subunits and was therefore, assigned to a dimer (Figure 27 A). Furthermore, the mass fractions of NSP9 monomers and dimers (AVG±SD, N=3; 75±11% and 25±2%, respectively) were determined, based on the intensity ratio of the assigned peaks. The low abundance of NSP9 dimer is consistent with a low binding affinity, as frequently reported in literature.

The NSP7 homo-dimer was assigned to two peaks (2374 *m/z*, 2670 *m/z*), but as a rule of thumb, three peaks are required to safely assign a mass species. Therefore, further validation was

needed, but no CID spectra were acquired. Nevertheless, by examining the *m/z* sodium adducts, it was found that molecular ions of the NSP7 dimer were present and therefore dimeric assignment was confirmed (Figure 27 B). Mass fractions were determined from intensities of clearly assigned peaks for NSP7 monomer and dimer (AVG±SD, N=3; 32±6% and 68±12%, respectively). In earlier reports, NSP7 was rarely described in the absence of NSP8. However, NSP7's ability of dimerization had already been noted, but no function has been suggested.

Some of the here presented results require further discussion. The found order of cleavage, demonstrated NSP8-9 to be cleaved relatively efficient, opposes the results from the FRET peptide assay (Section 4.2.1) where FPS8-9 was virtually inactive as a substrate. This deviation from full-length protein to peptide substrate could be due to missing structural environment in FPS, which merely represents amino acids of the cleavage site. In fact, two observations made by other researchers give indication for the reason. Fan et al. analyzed the secondary structure of NSP peptides and reported that NSP8-9, comprising the NNE motif and closely resembling the FSP8-9 tested here, has a higher ratio of alpha helices than other NSP peptides. Chuck et al. reported that β-sheets are required to promote efficient substrate binding. These studies indicate that an α-helix in FPS8-9 inhibits binding to M_{pro} and therefore cleavage.

After combining this information with the experimental results, it appears that the structural layout of the cleavage site is different from FPS8-9 to NSP8-9, and here, in fact, structural environment plays a role. Since NSP7-10 is a full-length and folded protein, the cleavage sites could be influenced by bending or stretching from the neighboring domains or even from structural elements further off. Therefore, processing results from the native-like substrate NSP7-10 appear more valid than from isolated FRET peptides.

The cleavage with full-length substrates as well as with peptides demonstrated that NSP9-10 is the most efficiently cleaved site. Particularly one study had suggested NSP9-10 in MHV to be a candidate as a long lasting precursor with a putative function *in vivo*. The results shown here do not confirm this for SARS-CoV, because NSP9-10 intermediate was never found upon cleavage, and further the determined order of cleavage indicated that such a long-lasting NSP9-10 intermediate would not be favorable.

The actual long lasting intermediate product was NSP7-8, most evident from native MS after 6 h of processing. Reduced cleavage leads to a covalently bound NSP7-8 intermediate, tailored from the polyprotein. Such an intermediate might be beneficial for the subsequent NSP7+8 complex formation. However, evidence remains elusive for this specific NSP7-8 tailoring because cleavage sites outside of NSP7-10 were not measured. Notably, some mass species such as NSP9-10, NSP8-9 or NSP8-10, could not be assigned to signals in the mass spectrum, most likely, because their generation by cleavage was unfavorable.

4.2.3 *Processing of SARS NSP7-9*

Processing was investigated with polyprotein NSP7-9-His. This substrate was shorter compared to His-NSP7-10, and processing lead to less products and thus a cleaner spectrum. The determined cleavage order of NSP7-9-His was consistent with the results described above.

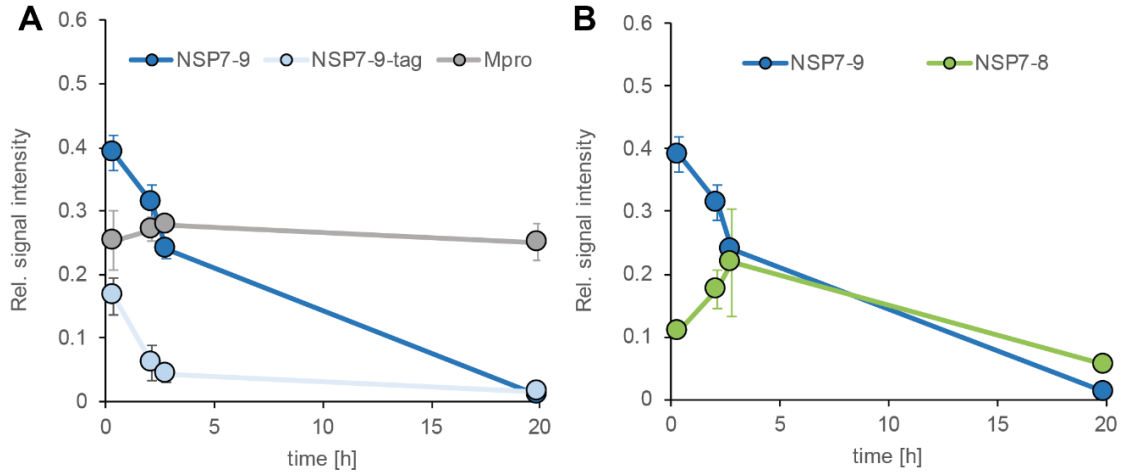


Figure 28: SARS NSP7-9-His processing in native MS: Signal over time. Shown in the graphs are relative signal intensities at different time-points. All data obtained in one experimental session and plotted in different graphs for the purpose of clear representation. **(A)** Initial substrate NSP7-9-His decreased quickly, due to efficient cleavage of the His tag. Relative intensity of protease remains stable. **(B)** NSP7-9 is converted in NSP7-8, as indicated from their diametrical relative intensities. Error bars indicate standard deviation (AVG±SD, N=3) 0.3±0.2 h; 2.1±0.1 h or the difference, when only two data points were available (AVG$|x_2-x_1|$, N=2) 2.7(0.4) h, 19.9 (0.2) h.*In vitro* processing and native MS analysis were carried out as described above. As expected, a time course of relative intensities showed different mass species dominate successively. In the beginning polyprotein and M_{pro}, then during processing intermediate products and at the end-point final products (Native mass spectra: Figure S 8).

The first step was conversion from NSP7-9-His to NSP7-9 (Figure 28 A), suggesting that the tag was cleaved more readily than the inter-domain junction. The protein construct NSP7-9-His had a C-terminal His-tag engineered via a cleavable linker (SKLQ///SGHHHHH), which represents an additional and artificial cleavage site. Thereby, a substrate was generated without any additional amino acids that could disturb cleavage efficiency or complex formation.

Second, M_{pro} converted NSP7-9 to the NSP7-8 intermediate (Figure 28 B). Similar as in the longer substrate, NSP7-8 was a lasting intermediate product, which was confirmed by CID (Figure S 9).

Converting NSP7-9 in NSP7-8 must have been facilitated through cleavage-based release of NSP9. In fact, the relative intensity of NSP9 became dominant in the mass spectrum, suggesting accumulation of NSP9 in parallel to generation of NSP7-8. (Figure 29 A).

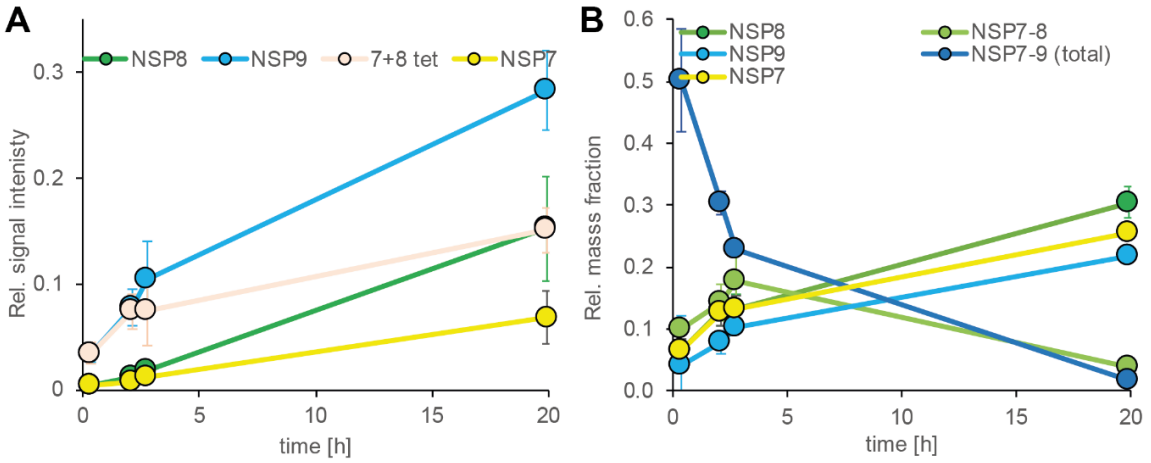


Figure 29: SARS NSP7-9-His processing in native MS: Signal over time, mass fraction of NSPs. Shown in the graphs are relative signal intensities at different time-points. All data obtained in one experimental session and plotted in different graphs for the purpose of clear representation. **(A)** Relative intensities of mass species as assigned in the spectra. NSP9 and NSP7+8 increase quickly and eventually NSP9 becomes the dominant signal in the spectrum. **(B)** Mass fraction, considering non-covalent complexes and showing the complete signal for the NSP7, NSP8 and NSP9 products. At the last time point (~20 h), equimolar presence is not reflected by the mass fractions. Error bars indicate according to data point (AVG±SD, N=3) 0.3±0.2 h; 2.1±0.1 h or (AVG($|x_2-x_1|$), N=2) 2.7(0.4) h, 19.9 (0.2) h.

At last, NSP7-8 was converted to NSP7 and NSP8 that in turn formed NSP7+8 complexes, which were identified in CID-MS (Complex formation of NSP7+8: Section 4.3.1).

Mass fractions (MFs) were calculated from the relative signal intensities (Figure 29 B), after the non-covalent complexes NSP7+8(2:2), NSP7+8(1:1), NSP7 dimer and NSP9 dimer were identified in CID-MS (Figure S 9). From the MFs, the order of cleavage did not become clear. However, also equimolar presence of NSP7, NSP8 and NSP9 was not reflected in the mass fractions at the last time point (~20 h). The reason for this could be that the MFs of NSP7 and NSP8 contained a considerable amount of NSP7+8 tetramer. The tetramer is a higher *m*/*z* ion (~65 kDa) that, most likely, had a better transmission in the Q-TOF compared to smaller NSPs (~10 to 22 kDa) and therefore gave unproportioned signal.

The findings from NSP7-9 processing by native MS were validated with SDS-PAGE as a complementary readout (Figure 30). The products detected at different time-points resembled the ones found in native MS analysis. Cleavage of the His-tag was clearly present after 1 h. During the same period, associated NSP9 and covalently bound NSP7-8 were detected, confirming the suggested order of cleavage. By all and large, analysis by SDS-PAGE ensured exclusive detection of covalent bound products, but lacked in mass resolution and sensitivity compared to native MS. For instance, M_{pro} bands overlapped with NSP7-8 precursor bands and hence, hindering their detection. Furthermore, protein staining was required, which emphasized bands of larger proteins, while bands of smaller proteins remained faint, such as NSP7 and NSP9.

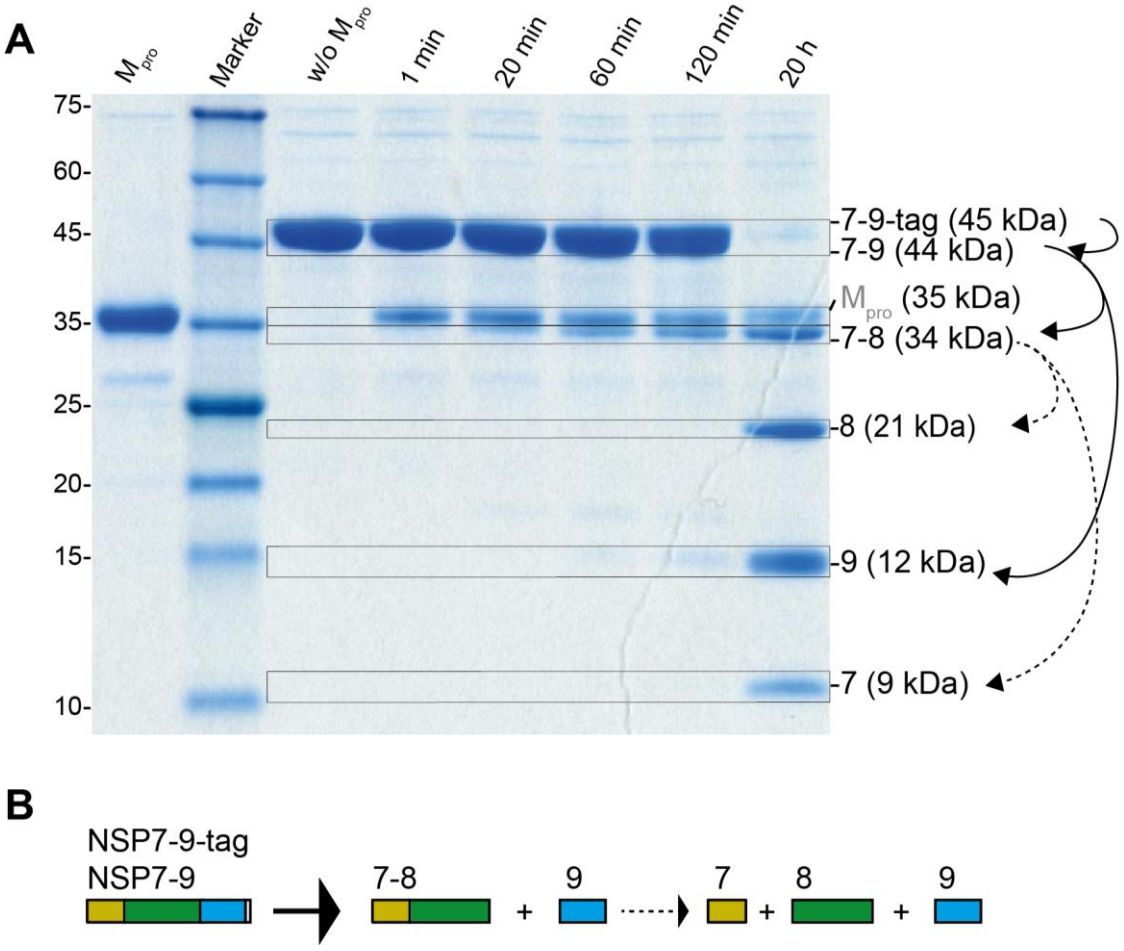


Figure 30: SDS-PAGE of SARS NSP7-9 processing. (A) For *in vitro* processing 3.2 µM SARS M_{pro} was incubated with 13 µM SARS NSP7-9-His (ratio ~1:4) at 4°C in 20mM phosphate buffer, 150 mM NaCl, 1 mM DTT at pH 8. SDS-PAGE was performed with a 4-12% gradient acrylamide Bis-tris gel with XT MES running buffer. Lane1, SARS M_{pro}; lane2, Roti Tricolor marker; lane 3, NSP7-9-His; lane 4-8, *in vitro* processing sample after indicated time points. Frames indicate for band assignment. Black arrows illustrate cleavage pathway as concluded from band shift. **(B)** Symbols illustrate conversion of NSP mass species, as concluded from the NSP7-9-His processing experiments.

4.2.4 Processing of SARS NSP7-8, NSP8-9 and NSP9-10

Relative efficiencies of M_{pro} at the SARS NSP7-10 sites was revealed with polyprotein substrates, but to determine absolute efficiencies, substrates with a single inter-domain cleavage site were required. Therefore, *in vitro* processing of full-length His-NSP7-8, NSP8-9-His, and NSP9-10-His was investigated.

In this experiment, cleavage reactions were started separately for the three samples and their products were analyzed by SDS-PAGE (Figure S 10) and native MS (Figure 31) at indicated time-points. To analyze, relative intensities were normalized to M_{pro}, of which concentrations were similar in the different reactions. In this experiment directly comparable intermediate products were absent and a cleavage order NSP7-8 to NSP8-9 was not as evident. However, the results suggested that the NSP9-10 junction is cleaved with highest relative efficiency, consistent with the data acquired from larger polyproteins (Processing of SARS NSP7-10: Section 4.2.2).

To continue, the experiments require optimization. Foremost, the sample constructs were not ideal because NSP8-9-His and NSP9-10-His had artificial cleavage sites between NSPs and His-tag. These sites introduced additional products, which hindered assignment of signal intensities in native MS. Substrates that are more suitable could be generated with mutations that render the artificial sites inactive. Furthermore, comparison and normalizing signals in between spectra turned out cumbersome. To solve this challenge, different samples should have been co-incubated. At last, determining higher efficiencies requires a higher rate of sampling over a longer time-period. While NSP7-8 and NSP8-9 processing requires sampling for hours, the NSP9-10 processing requires sampling within minutes.

One additional observation revealed details about substrate specificity. Both artificial cleavage sites of NSP8-9-His and NSP9-10-His had identical His-tags leading to NSP9-His (ATVRLQ///ARGGHHHHH) and NSP10-His (REPLMQ///ARGSGHHHHH), respectively. Native mass spectra showed that in NSP8-9-His the His-tag was cleaved before, while in NSP9-10 it was cleaved later than the inter domain junction. The different efficiencies must be due to the C-terminal amino acids of NSP proteins that represented the only difference between both substrates since the tag was the same. Whether NSP10 is also poorly separated from the NSP11/12 domain in the context of pp1a/ab processing remains to be investigated.

After studying the NSP7-10 processing of SARS-CoV, it was intriguing to bring the results in a broader context by comparing them to other coronavirus species.

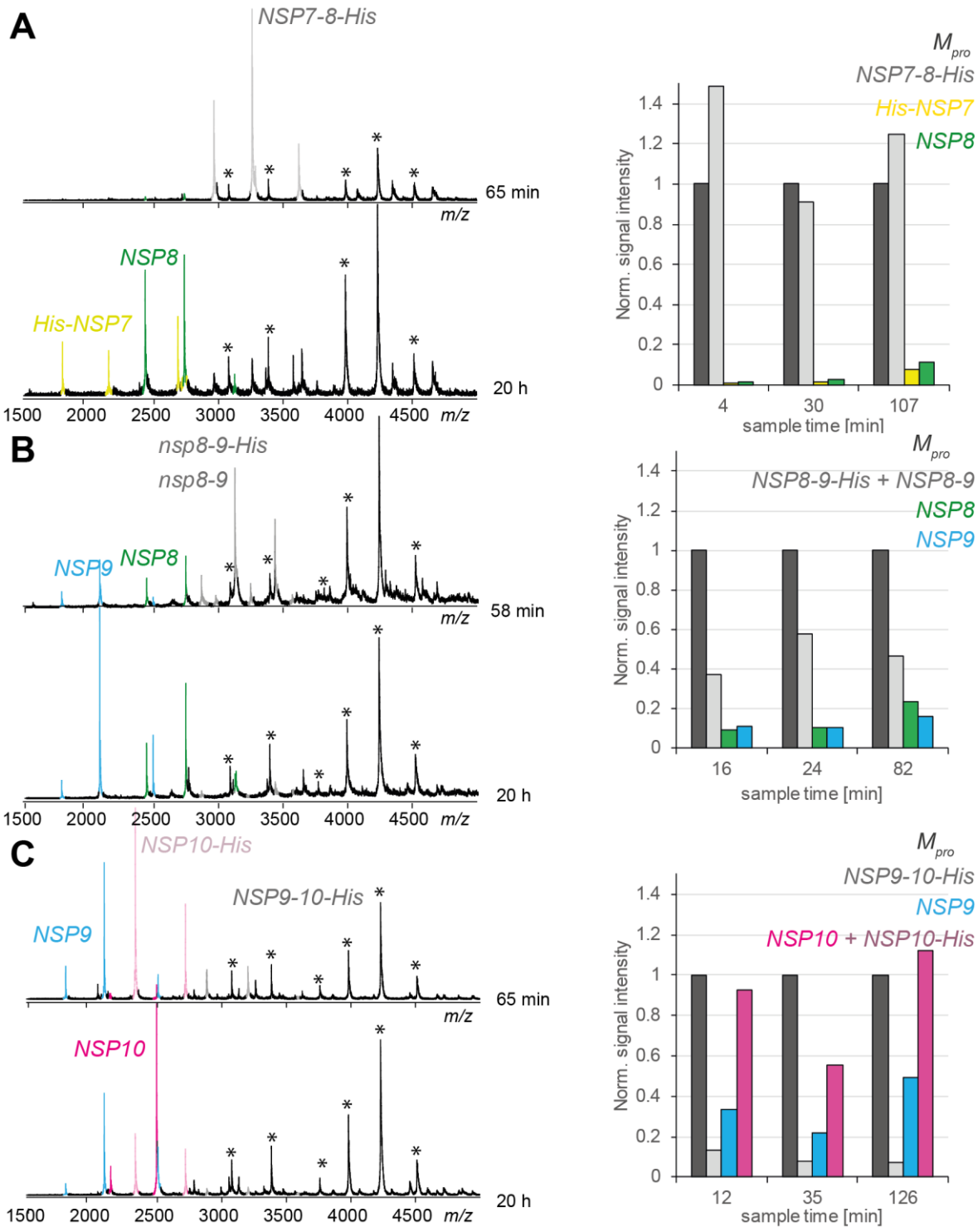


Figure 31: Native MS analysis of SARS NSP7-8, NSP8-9 and NSP9-10 as substrates. For in *in vitro* processing, 3 µM M_{pro} was incubated with 14 µM of substrate (ratio ~1:5) at 4°C in 250 mM AmAc, 1 mM DTT at pH 8. Shown are exemplary native mass spectra of processing at the indicated time points. **(A)** Substrate His-NSP7-8 and products His-NSP7 and NSP8, **(B)** Substrate NSP8-9-His and products NSP8, NSP9 and NSP9-His. **(C)** Substrate NSP9-10-His and products NSP9, NSP10 and NSP10-His. Bar diagrams plotted from relative intensities and normalized for M_{pro}. Tagged and untagged proteins were summed up, to illustrate the ratio of protein before and after cleavage at the inter-domain junction. NSP10 + NSP10-His for cleaved; NSP8-9 + NSP8-9-His for uncleaved. Bar diagrams were chosen instead of a time course in a diagram because there were not enough data points to confidently show a trend.

4.2.5 *Processing of 229e NSP7-10*

The analysis was extended to proteins of hCoV 229e. Matter of investigation was the NSP7-10 region, having 47% sequence identity to SARS and virtually conserved cleavage sites (Figure S 15).

For *in vitro* processing, a 229e NSP7-10-His with a non-cleavable tag was incubated with its natural protease 229e M_{pro}. A similar approach as described above was used, which combined *in vitro* processing followed by SDS-PAGE and native MS (Processing of SARS NSP7-10: Section 4.2.2). In the course of processing, different categories of mass species were dominant in the native mass spectra; at the beginning, substrate, in mid-reaction, intermediate products and complexes thereof and at the end, NSP monomers (Figure 32).

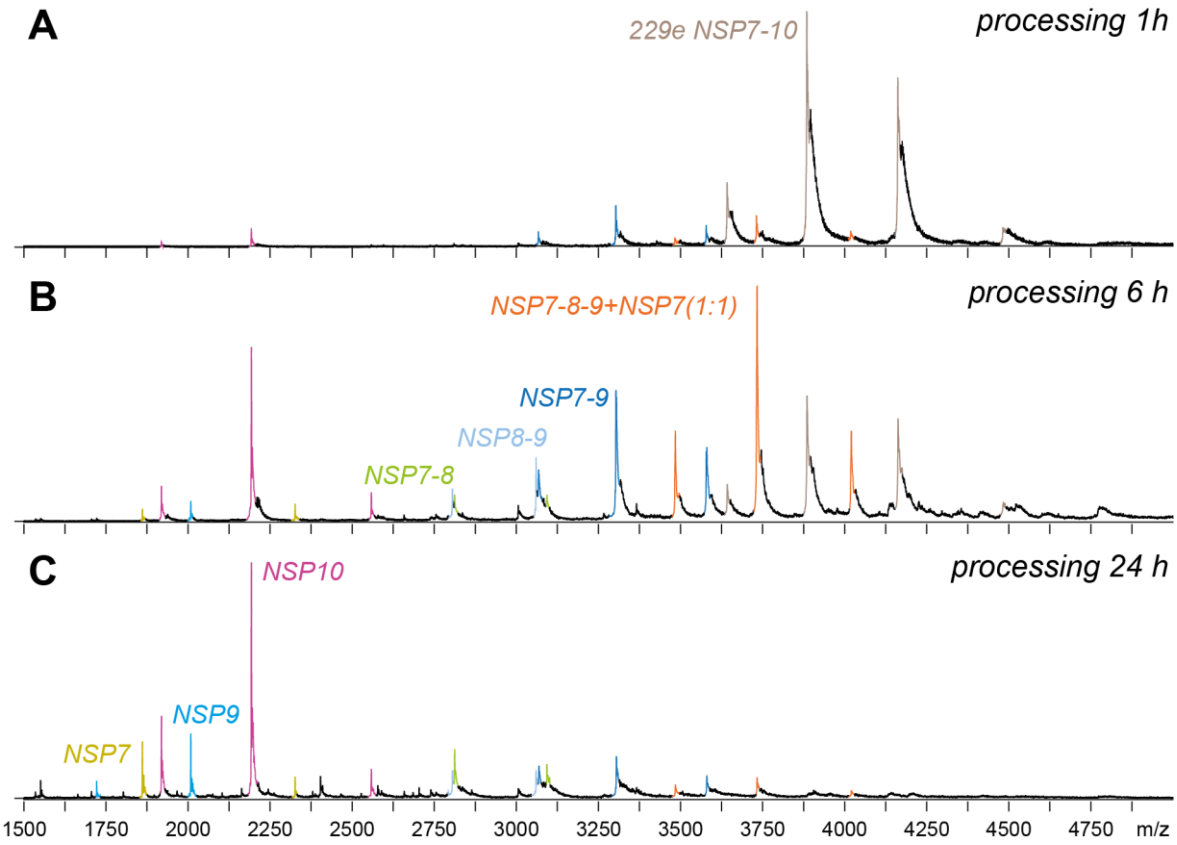


Figure 32: Native MS dynamics of 229e NSP7-10 processing: Exemplary spectra. Exemplary native mass spectra illustrate processing. For *in vitro* processing, 1.25 µM 229e-CoV M_{pro} and 12.5 µM 229e NSP7-10-His (~ratio 1:10) were incubated at 4°C in 250 mM AmAc, 1 mM DTT at pH8. Spectra acquired at indicated reaction time. **(A)** Mass spectrum at 1 h showing dominant peak series for substrate NSP7-10. **(B)** Mass spectrum at 6 h with dominant peaks assigned to intermediates NSP7-9 (dark blue) and NSP7-9+NSP7 complex (orange). **(C)** Mass spectrum at 24 h showing dominant peaks for NSP monomers, NSP7 (yellow), NSP9 (blue) and NSP10 (pink) (**Table S 1: Molecular weights (MWs) of NSPs, their cleavage products and complexes.**).

The order of several reactions became evident from monitoring relative intensities (Figure 33). From the diametric change of signal intensity of substrate and products, it became clear that NSP7-10 was depleted while NSP products were generated (Figure 33 A). The signals for 229e

M_{pro} could not be monitored, due to heterogeneity of this protein (Native mass spectra of proteases: Figure S 5).

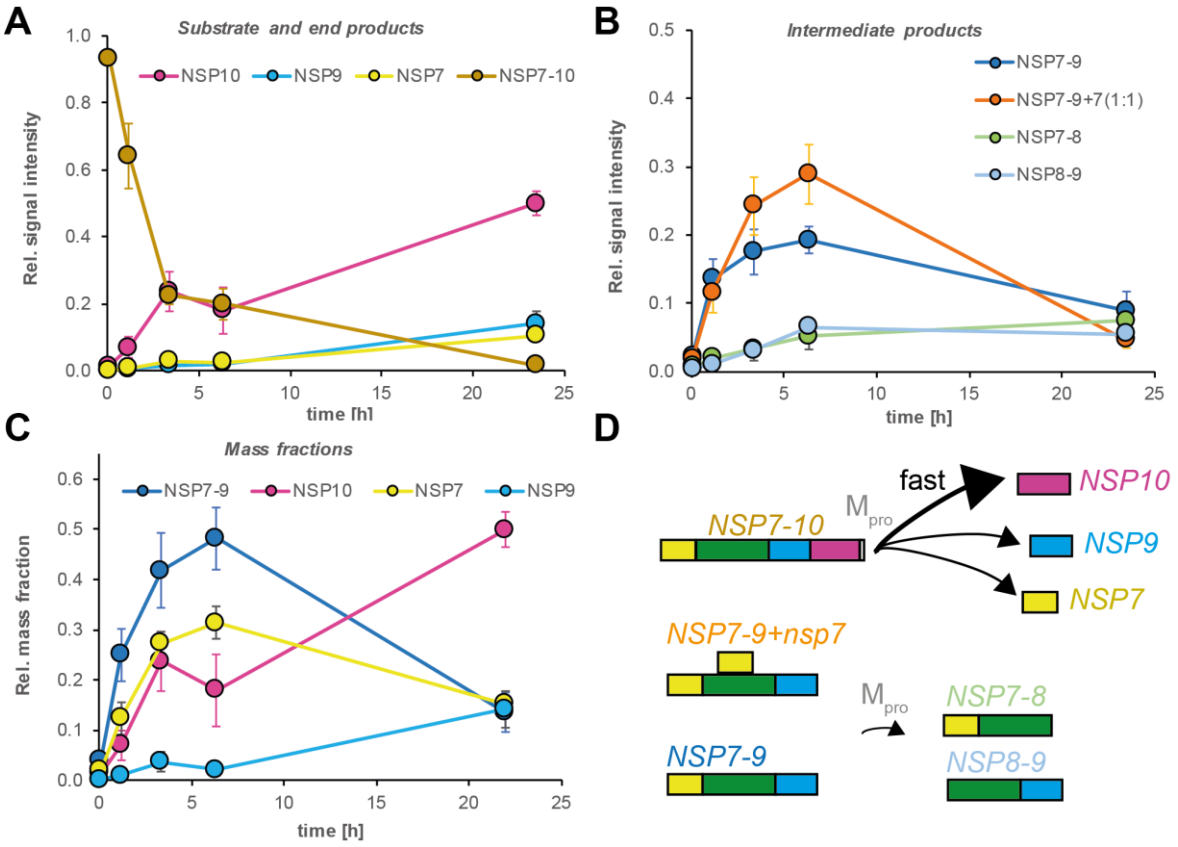


Figure 33: 229e NSP7-10 processing in native MS: Signal over time. Time course of 229e NSP7-10-tag *in vitro* processing. Shown in the graphs are relative signal intensities at different time points. All data were obtained in one experimental session and plotted in different graphs for the purpose of clear representation. **(A)** Relative intensity of NSP7-10 substrate decreases and in parallel, NSP products increase. NSP10 becomes the dominant mass species. **(B)** Relative intensity of cleavage intermediates. Increasing and then decreasing NSP7-8 (grass green), NSP8-9 (blue), NSP7-9 (light blue) and NSP7-9+NSP7 protein complex (orange) showing the temporal presence of these mass species. **(C)** Relative mass fractions showing extensive generation of NSP7-9 in the beginning of the reaction, but eventually cleavage of this intermediate into NSP monomers. The NSP7 and NSP10 showing similar increase that is higher than NSP9 increase, indicating their cleavage efficiency off from the NSP7-10 substrate. Error bars depict standard deviation (N=3). Time points (AVG±SD, N=3): 1.2±0.2 h, 3.4±0.2 h, 6.3±0.3 h and 23.5±0.5 h. **(D)** Symbols illustrate evolution of NSP mass species, as concluded from the 229e NSP7-10 processing experiments.

The increase of relative intensities for NSPs was prominent for NSP10, while NSP7 and NSP9 only increased slowly and NSP8 was not detected at all. The high increase rate of NSP10 led to the conclusion that the NSP9-10 site had the relative highest substrate efficiency, similar as determined for SARS NSP7-10.

Relative intensity of intermediate products revealed order of cleavage of other sites (Figure 33 B). Most abundantly and earliest present was NSP7-9, the counter-product of NSP10 release, and NSP7-9+NSP7(1:1), a protein complex which had not been described before. Formation of

the latter, required NSP7-9 and monomeric NSP7, which must have been generated by cleavage of the NSP7-8 site. However, both intermediates contained NSP7-9, and therefore this finding supported the conclusion, that the NSP9-10 site is cleaved readily.

Relative intensities of the intermediate products NSP7-8 and NSP8-9 increased concordantly with the decline of NSP7-9 and NSP7-9+NSP7(1:1). This course of signal demonstrates the order of generated mass species, particularly from NSP7-9 and NSP7-9+NSP7(1:1) to NSP7-8 and NSP8-9 (Figure 33 D).

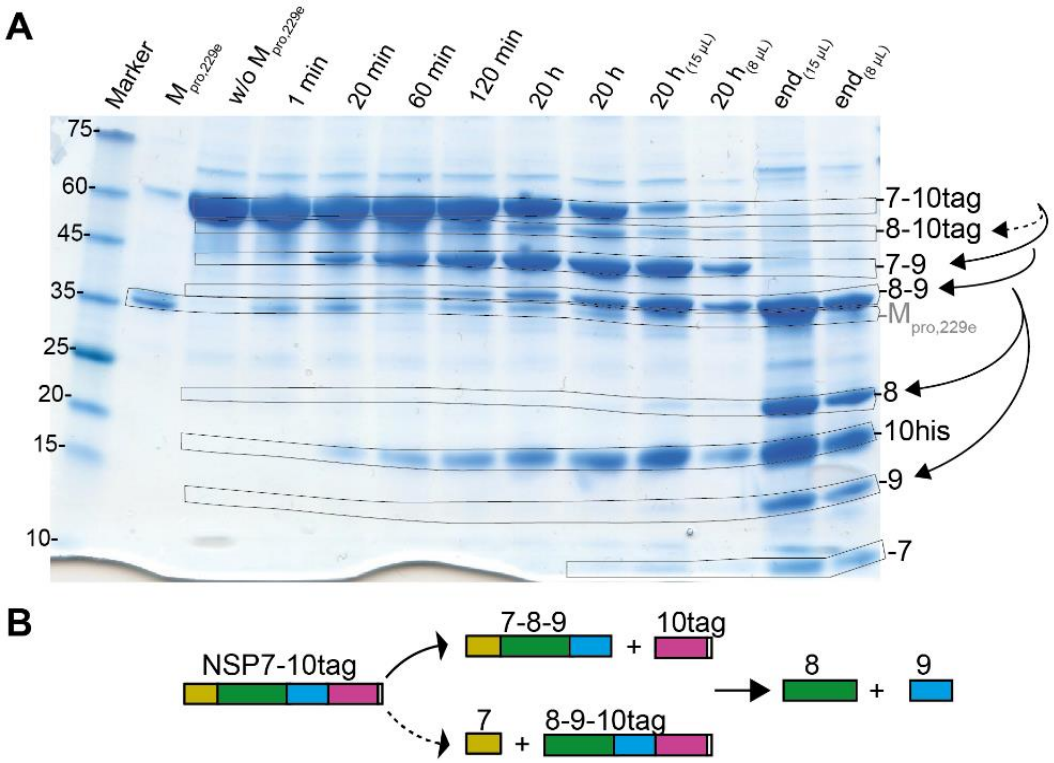


Figure 34:SDS-PAGE of 229e NSP7-10 processing. For *in vitro* processing 229e NSP7-10-His was incubated with 229e M_{pro} at 4°C in 20mM phosphate buffer, 150 mM NaCl, 1 mM DTT at pH 8. SDS-PAGE performed with a 4-12% gradient acrylamide Bis-tris gel with XT MES buffer. Lane1, molecular weight marker; lane2 229e M_{pro}; lane 3, 229e 7-10-His; lane 4-10, *in vitro* processing sample after dedicated time points. Frames (grey) indicate for band assignment. Black arrows illustrate cleavage pathway suggested by band shift. Symbols below illustratively describe course of NSP mass species, as concluded from the data.

To simplify, mass fractions (MFs) were calculated. Here, the mass fractions of NSP7-9 and NSP7 each contain the relative intensity of NSP7-9+NSP7, which also represents their cleavage off of the polyprotein (Equations for MFs: 3.8). The MFs indicated clearly that at the beginning NSP7-9 was the prevailing intermediate generated from NSP7-10. However, the MFs of NSP10 and NSP7 first increased similarly, which suggests near equal cleavage order of NSP7-8 site and NSP9-10 site. Nevertheless, at the end, the MF of NSP10 was much higher than that of NSP7. In fact, intermediate products NSP7-8 and NSP8-9 were detected at the last time point (~24 h)

indicating incomplete processing. This must have resulted in a non-equimolar presence between the single NSPs and therefore, the uneven MFs.

At last, judging from the high increase of NSP7 and NSP10, which was never reached by NSP9, it appears that the NSP8-9 site was cleaved with the lowest relative efficiency. Unexpectedly, NSP8 could never be detected, nevertheless, presence of NSP7-8 and NSP9 suggests that NSP8-9 was cleaved, but probably slower than the other two cleavage sites. After all, processing of 229e NSP7-10 was also determined by SDS-PAGE (Figure 34 A). Even though SDS-PAGE had a lower resolution than native MS, the results agreed and the generation of the mass species during processing became evident (Figure 34 B).

Some noticeable observations upon 229e NSP7-10 processing experiments require a brief discussion. Surprising, there was a hetero-dimeric interaction between NSP7-9 and NSP7, forming the protein complex NSP7-9+NSP7(1:1). CID proved its non-covalent nature (Figure S 6). Such a complex with uncleaved regions of the polyprotein has not yet been described. Most probably, this interaction is facilitated by NSP7 and NSP8, as implicated by the binding of these well-studied domains from other CoV species, but not yet from 229e CoV. Remarkably, one domain of NSP8 would be able to bind two NSP7, one covalently and one non-covalently. However, exact assembly of this complex remains elusive.

Low intensity or absence of mass species hindered semi-quantitative analysis of 229e NSP7-10 processing in native MS. Particularly NSP8 and intermediate products that contain NSP8 domain were less intense or even not detected at all. Noticeably, if NSP7 was linked covalently to NSP8, as in mass species such as NSP7-8 or NSP7-9, then they were observed at the expected intensities. However, if NSP7 was cleaved off from NSP8, as in NSP8, NSP8-9 or NSP8-10, these mass species were absent or of lower intensity as their cleavage counter-products would suggest. A reason for this could be that structural homologues of NSP8 have an elongated N-terminal domain, which is typically aggregation prone. The low intensity of some products could be explained when NSP8 aggregated or was unable to ionize when its N-terminus had lost protection by NSP7. Furthermore, an incorrectly folded NSP8 could have changed the observed order of products in this experiment.

At last, SDS-PAGE was carried out as well and has some advantages over native MS in processing experiments due to detection of strictly non-covalent products regardless of ionization efficiencies and complex formation. However, care has to be taken when interpreting these results because, as presented above, NSP8-9 might be extremely prone to aggregation and reside in clusters, where M_{pro} cleavage is unlikely. Therefore, native MS and SDS-PAGE could be both similarly biased and the cleavage of NSP8-9 might not be as inefficient, as indicated from the results.

4.2.6 *Processing of FIP NSP7-9*

To complete the presentation of polyprotein processing one specific native MS experiment is included with proteins of FIP CoV. The FIP NSP7-10 region has sequence identity of 46% with SARS and of 45 % with 229e, but contains conserved cleavage sites. The protein substrate available encompasses merely NSP7-9 and was therefore shorter than the above described substrates. However, presence of the NSP7-8 site and the non-canonical NSP8-9 site allowed comparison of their relative cleavage efficiency. In the following paragraphs, a native MS analysis from two time-points of processing is presented.

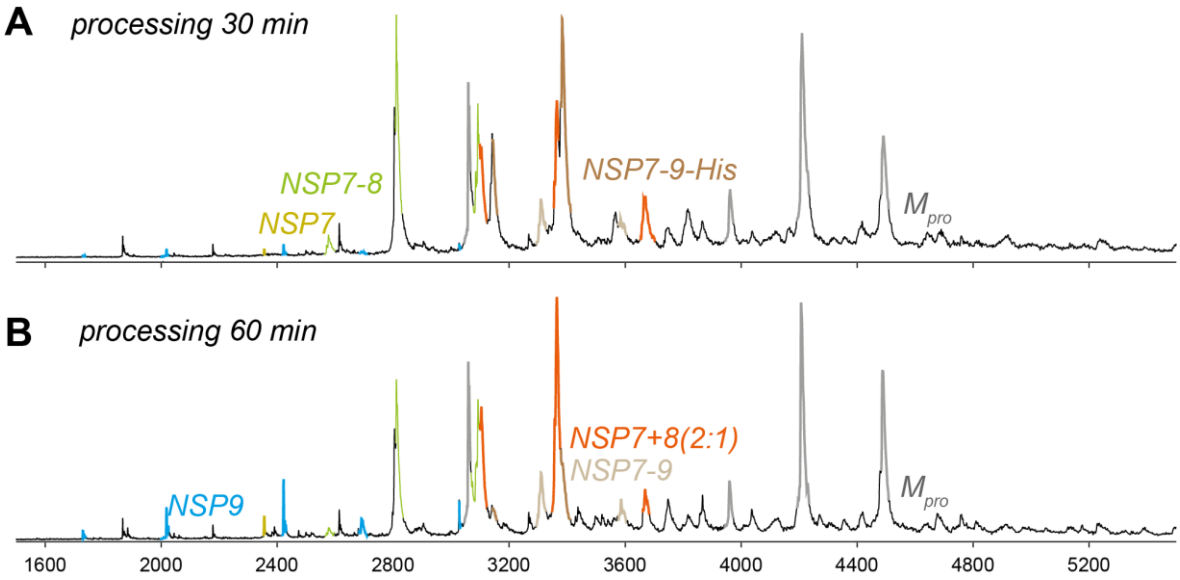


Figure 35: Native MS dynamics of FIP NSP7-9 processing: Exemplary spectra. Exemplary native mass spectra illustrate processing between two time points. For *in vitro* processing 5 µM FIP M_{pro} and 20 µM FIP NSP7-9 (ratio ~1:4) were incubated at 4°C in 250 mM AmAc, 1 mM DTT at pH8. After mixing the components, samples were injected into an electrospray capillary and native MS spectra were acquired after indicated time points. **(A)** Mass spectrum at 30 min showing dominant signals for NSP7-9-His and intermediate product NSP7-8 and M_{pro}. **(B)** Mass spectrum at 60 min showing dominant signals for NSP7+8(2:1) complex (orange) and M_{pro}(grey). Additionally, NSP7 (yellow) and NSP9 (blue) mass species of lower intensity were assigned. Some peaks could not be assigned to any NSP product or combination thereof, and are results either from in solution fragmentation or molecular crowding due to high sample concentration. (**Table S 1: Molecular weights (MWs) of NSPs, their cleavage products and complexes.**)

For *in vitro* processing a FIP NSP7-9-His with a cleavable tag was incubated with FIP M_{pro}, then samples were withdrawn after two time-points and native mass spectra recorded (Figure 35). For data analysis, peak signals were assigned to processing products by their theoretical molecular weight (Figure 36). No statistical significance could be reached from two measurements, however, the order of several reactions during NSP7-9 processing became evident from comparing the products relative intensities.

At 30 min, which was the first time-point of sampling, the main substrate had already been subject to extensive processing and was converted into other mass species. A steep decrease of NSP7-9-His showed conversion into NSP7-9 upon cleavage of the His-tag. Remarkably, the NSP9 domain and the tag were not linked by additional amino acids (ATVRLQ///HHHHHH). This reaction highlighted the importance of the N-terminal amino acids (P6-P1) of the cleavage site, as shown by the efficiency of cleavage of artificial sites in SARS NSPs (Section 4.2.4).

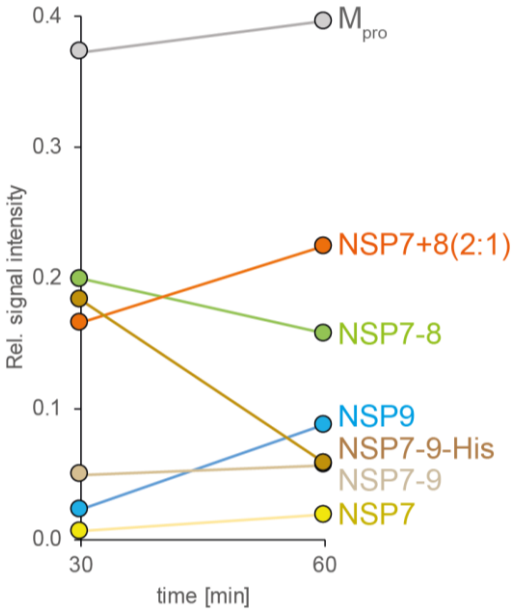


Figure 36: FIP NSP7-9 processing by native MS: Signal over time. Relative signal intensity of assigned mass species in the native MS spectra at 30 min and at 60 min (Figure 35). Lines between the data points indicate increase or decrease of relative intensity.

Most important were the relative intensities for the relative cleavage efficiency between NSP7-8 and NSP8-9. Judging from the relative signal intensities, NSP7-9 remained rather stable but NSP9 increased from 30 min to 60 min. The NSP7-9 signal stability was interpreted as a quasi-steady state resulting from constant generation and depletion. On the one hand, cleavage of the His-tag generated NSP7-9, on the other hand NSP7-9 was depleted by cleavage of its junction NSP8-9. Further evidence for this reaction was the high relative intensity of intermediate NSP7-8. Typically for longer lasting intermediate products, there is first an increase and then decrease in their relative signal intensity. Increasing signal for NSP7-8 was never observed, most likely, the 30 min time point was already too late and the maximum concentration of NSP7-8 had already been passed.

The results suggest the NSP8-9 site is cleaved equally or even more efficiently than the NSP7-8 site, as the results above similarly indicated for SARS. To confirm this, further measurements are needed with a longer NSP7-10 substrate.

In a following chapter, the conclusions drawn from polyprotein processing are also discussed in context of the virological relevance (Chapter 5: Conclusion and outlook).

4.3 Protein complex formation of NSP7+8

Upon processing of NSP7-10, I frequently observed interacting subunits (Section 4.2.2.). Excited for more detailed studies, I focused on the NSP7+8 complexes since they had been found as functional building blocks of the CoV polymerase and therefore, were of particular interest. Here, in a second set of experiments, I analyzed their complex formation by native MS and determined their complex stoichiometry by tandem MS. In this chapter, I present and interpret the results of NSP7+8 complex formation from several coronavirus species, namely SARS, TGEV, FIP, PEDV and 229e.

4.3.1 Complexes of SARS NSP7+8

Initially, I tested the complex formation of SARS NSP7+8 with protein samples from the processing experiments. In the following paragraphs, I describe first native MS of SARS complexes, then their collision-induced dissociation, and lastly a comparison of the obtained data with available high-resolution structures.

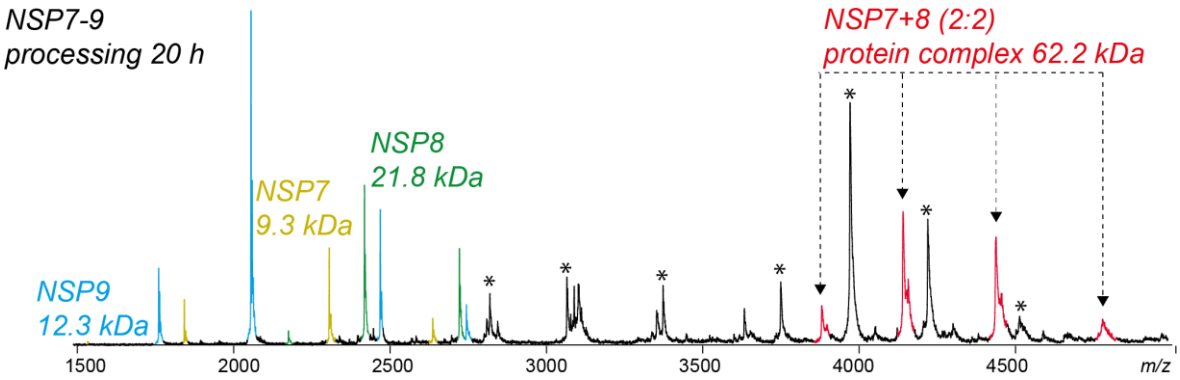


Figure 37: SARS NSP7+8 interacts as a (2:2) hetero-tetramer. Native mass spectrum of NSP7+8 complexes formed after cleavage of NSP7-9. For *in vitro* processing, 2 µM SARS M_{pro} was incubated with 14 µM SARS NSP7-9-His (ratio ~1:7) at 4°C in 300 mM AmAc, 1 mM DTT at pH8 overnight. Highlighted is the newly emerging putative NSP7+8 (2:2) tetramer (red). Complex assignment confirmed in CID product ion spectra (Figure 38). Other assigned peaks correspond to M_{pro} (asterisk) and the monomer processing products NSP7 (yellow), NSP8 (green) and NSP9 (blue). (**Table S 1: Molecular weights (MWs) of NSPs, their cleavage products and complexes.**)

To induce efficient complex formation of NSP7 and NSP8, these proteins were required to contain only the authentic amino acid sequence. To achieve this, polyproteins were cleaved by the coronaviral protease M_{pro} that considered their original cleavage sites and thus, tailored them to possess authentic termini.

To specifically investigate SARS NSP7+8 complex formation, NSP7-9-His was processed *in vitro* and complex formation sampled via native MS (Figure 37). As expected, peak envelopes were detected for the monomeric products NSP7, NSP8 and NSP9 and for M_{pro} monomer and dimer. Most importantly, one series of intense peaks was assigned to a hetero-tetramer of NSP7+8(2:2) (62.2 kDa). Based on the MWs of NSP7 and NSP8, the stoichiometry could be predicted (2x9.3 kDa + 2x21.8 kDa = 62.2 kDa). For further evidence, tandem MS was performed.

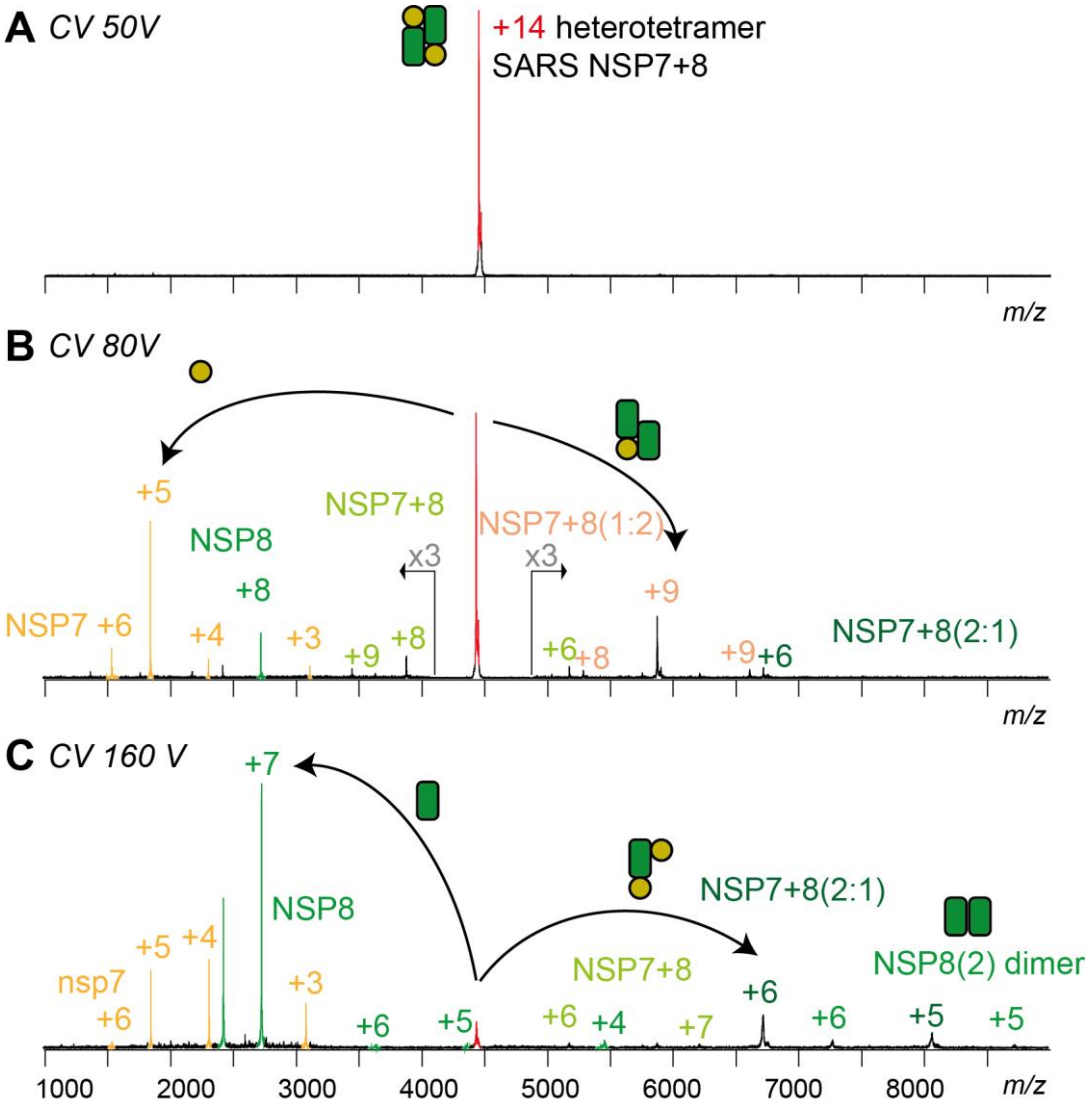


Figure 38: Collision-induced dissociation of SARS NSP7+8 hetero-tetramer. Dissociation triggered in Q-TOF tandem MS with increasing CV at 14 µbar argon. Arrows depict matching dissociation products. **(A)** Precursor ion spectrum at CV 50 V showing high precision of precursor ion selection (+14, 4449 *m/z*). **(B)** Product ion spectrum at CV 80 V showing two pathways of dissociation, preferably into NSP7 and NSP7+8(1:2) trimer or into NSP8 and NSP7:8(2:1) trimer. Magnification of low and high *m/z* range as indicated. **(C)** Product ion spectrum at CV 80 V showing that high charged NSP8 (8+) becomes the most intense peak. Additionally, product ions NSP7+8(1:2) and NSP8(2) dimer, generated by follow up dissociation of the trimers, provide evidence for direct interactions within the complex.

For tandem MS, three different charge states of the putative NSP7+8 hetero-tetramer (4449 *m*/*z*, +14; 4153 *m*/*z*, +15; 2893 *m*/*z*, +16) (Figure 38 A; Figure S 6) were selected as precursor ions in the quadrupole and were subjected to the collision cell. To induce collision-induced dissociation (CID), activation of the ions was triggered by carefully increasing the collision voltage (CV).

Upon CID, the overall dissociation pattern was revealed (Figure 38 B). Dissociation confirmed the complex stoichiometry and provided evidence for assembly of four individual subunits into NSP7+8(2:2). Furthermore, product ions NSP7+8(1:1) and NSP8(2) dimers provided evidence for interfaces within the complex.

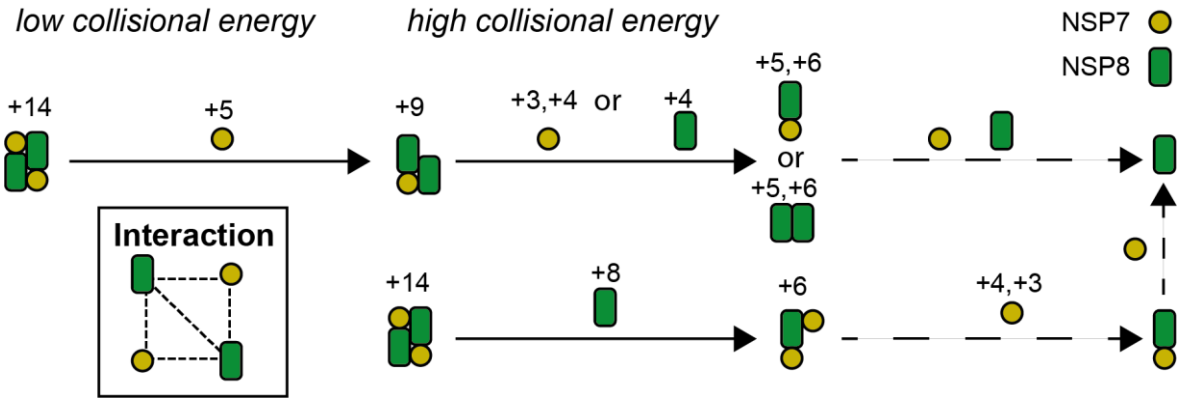
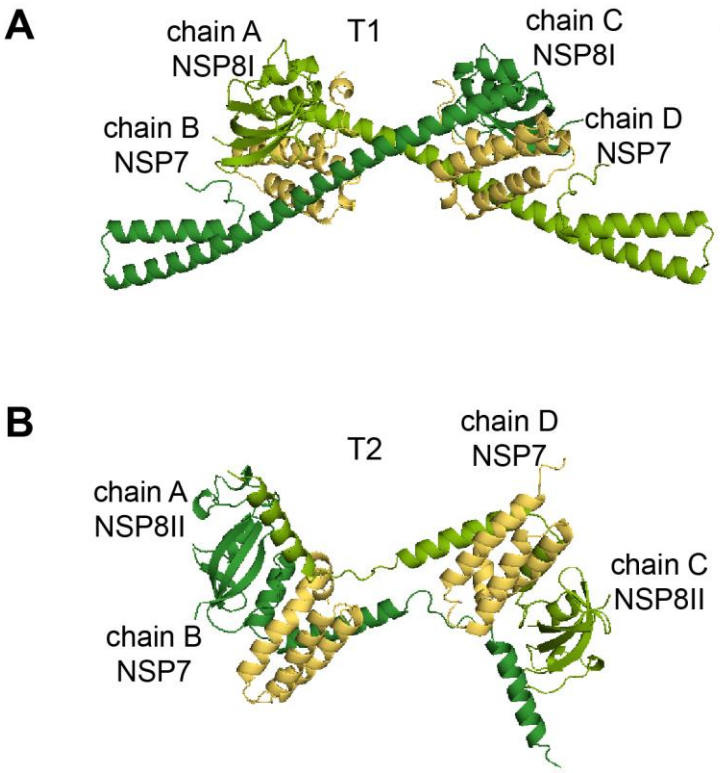


Figure 39: Gas-phase dissociation pathways of NSP7+8(2:2). Two alternative dissociation pathways of SARS tetramer and interaction of subunits within the tetramer as concluded from CID spectra. **Main pathway (top):** Initially, NSP7+8(2:2) dissociates into NSP7 and NSP7+8(1:2), suggesting a peripheral positioning of NSP7 within the complex. In a follow up-dissociation NSP7+8(1:1) is detected at elevated collisional energy. **Alternative pathway (bottom):** At higher collisional energy, dissociation into NSP8 and NSP7+8(2:1) is preferred. NSP8 dimers also appear. **Interaction map:** Subunit interactions as concluded from CID-MS results. Symbols illustrate molecular ions from different mass species found as dissociating ions in CID. Black arrows indicate for the ejected ions. Dashed arrows indicate for unobserved dissociations that most likely follow up. Charge states labelled taken from dissociating ions found in CID SARS (+14) and when combined match the charge state of the precursor ion.

Two alternative pathways of dissociation revealed location of subunits within the complex (Figure 39). At lower collisional energy, hetero-tetramer dissociated into NSP7 and NSP7+8(1:2), and alternatively, at elevated collisional activation, into NSP8 and NSP7+8(2:1). The alternative dissociation pathways resulted from two analogue routes of energy absorption, possibly due to similar binding contribution of NSP7 and NSP8 within the complex or two different gas phase conformers. Since both pathways were observed in parallel, the presence of two NSP7+8(2:2) conformers is further discussed.

The above described SARS hetero-tetramer NSP7+8(2:2) was partially in line with results published elsewhere (Figure 40 A and B). So far, molecular structures of SARS NSP7+8 had been reported in two oligomer conformations, as hexa-decamer(8:8) of wildtype and as a dimer (1:1) with N-terminally chopped NSP8 [86, 87]. Even though hexa-decameric(8:8) oligomerization

[86] was not observed in native MS, this model is better suited for further analysis due to its full-length sequence.



C

Interface	Buried surface (Å^2)	Buried surface (%)
T1 total	10281	100
NSP7 in T1	2153	21
NSP8I in T1	2988	29
NSP8I:NSP8I	835	8
T2 total	10377	100
NSP7 in T2	1988	19
NSP8II in T2	3201	31
NSP8II:NSP8II	1214	12

Figure 40: Structural candidate for NSP7+8(2:2). Molecular structures that agree with stoichiometry and subunit interaction as indicated by native MS and tandem MS. Hetero-tetramer sub-complexes **(A)** T1 and **(B)** T2 extracted (PyMol) from a hexa-decameric SARS NSP7+8 as solved by X-ray crystallography and available at PDB 2G9T [86]. **(C)** Shown in the table are selected buried surface areas as an approximation for binding interface. To get approximation of buried interface, the surface area of single subunits was subtracted by the surface area of T1 or T2 with PyMol and get_area input order. Here, NSP7/NSP8 in T1/2 is the buried surface between a single subunit and the complex. NSP8I:NSP8I/NSP8II:NSP8II is the buried surface by the NSP8 scaffold alone.

Within the hexa-decamer, there are two different hetero-tetramer subunits, T1 and T2. They structurally differ in folding of the N-terminal alpha helix of NSP8. However in both, the NSP8 builds a scaffold, in which the C-terminal head of one molecule interacts with the N-terminal tail of the other molecule. NSP7 subunits bind to the NSP8 molecules in a sandwiched position, but do not interact with each other. These molecular architectures of T1 and T2 are both in agreement with the CID results.

Analyzing the binding interfaces of T1 and T2, the structures were further related to CID results. The interfaces of NSP8(2) dimer scaffold as well as of NSP7 and NSP8 are mainly hydrophobic. However, three polar bonds within or at proximity of NSP8 shaft domain facilitate binding to NSP7 (Figure 54). Since polar bonds are more likely to persist in gas phase than hydrophobic bonds, it was not surprising to find rather remainder ions of NSP7+8(1:1) than of NSP8(2) dimer.

More information was revealed by calculating the buried surface areas *in silico*, as an approximation for total binding interface in T1 and T2 (Figure 40 C). Their total buried surface areas were found to be virtually equal, but the NSP8(2) dimer scaffold alone was tighter in T2 because a loop in the NSP8 N-terminus allows it to bend back and interact tighter with the other NSP8. In T1, the NSP8 scaffold has a different fold. Its N-terminal helix is not bent but elongated and interacts with another tetramer in the crystal lattice. A structure like T1 appears unlikely in the absence of a crystal environment. Hence, it was concluded that T2 represents a more likely structure of NSP7+8 in solution.

Altogether, the proteins SARS-CoV NSP7 and NSP8 specifically interacted as a NSP7+8(2:2) heterotetramer protein complex. The results indicate that within this complex, two NSP8 have an interface and additionally bind to NSP7. These findings agreed with hetero-tetrameric sub-complex T2 of an available high-resolution structure. To continue the experiments, analysis was extended to other homologue protein complexes, which had been described in different stoichiometry.

4.3.2 *Complexes of NSP7+8 from TGEV and FIP*

FIP and TGEV have homologue NSP7-8 domains with a sequence identity of 93.9 % (Clustal Omega, Figure S 4). Complex formation of FIP but not TGEV NSP7+8 has already been reported [84]. This proved to be an excellent opportunity to further test the native MS approach for complex analysis and compare the two complexes regarding their similarities and differences. In the following paragraphs, I present results about first, native MS of both protein homologues, then CID results for TGEV and FIP, and at last homology model of TGEV considering the obtained information.

Initially, the precursor samples TGEV and FIP NSP7-8-SGSG-His were detected as monomers (Protein constructs listed in Table 3). Then, upon M_{pro} cleavage of the precursor, NSP products of both viral species behaved remarkable similar regarding non-covalent interactions.

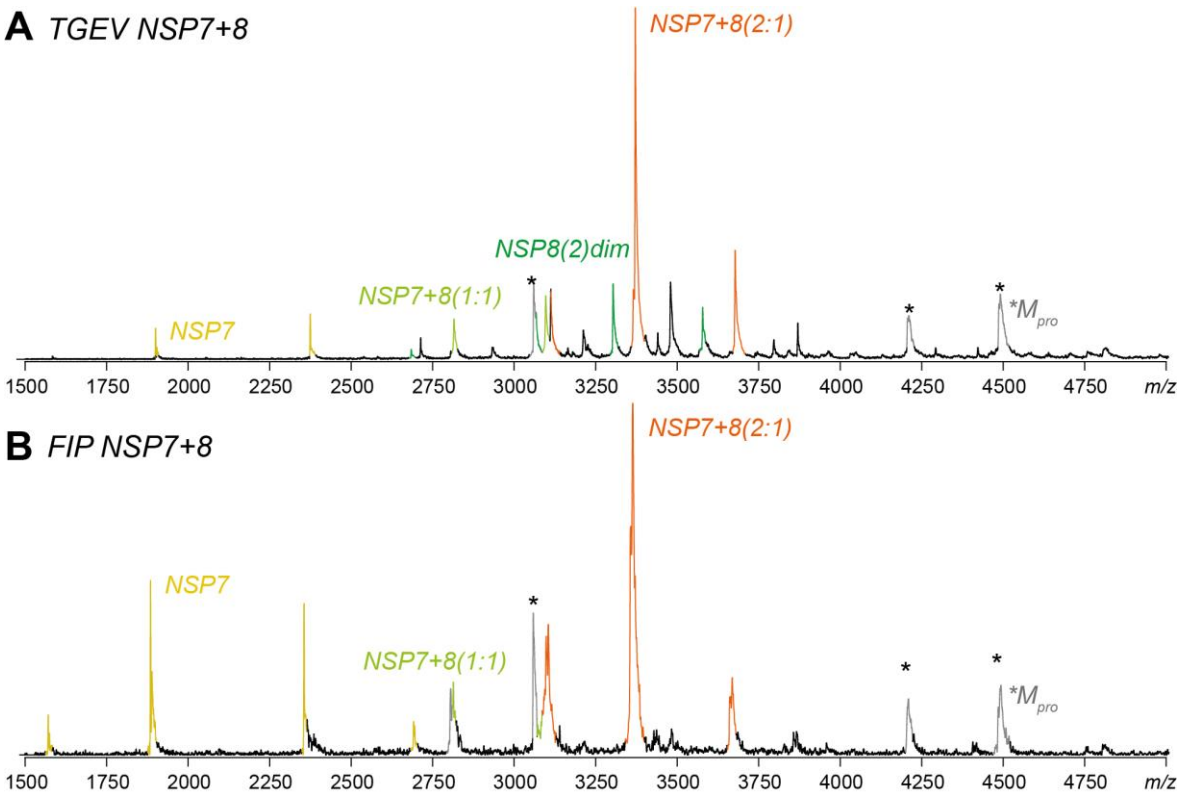


Figure 41: FIP and TGEV NSP7+8 form a hetero-trimer. Native mass spectra of **(A)** TGEV and **(B)** FIP NSP7+8 complexes after cleavage of NSP7-8. For cleavage, SARS M_{pro} was separately incubated with 17.5 µM FIP and 15 µM TGEV NSP7-8-NNE-His with a cleavable N-terminal His-tag (ratio ~1:10) at 4°C in 300 mM AmAc, 1 mM DTT at pH8 overnight. Complex formation is not influenced by the non-natural protease, which has virtually similar substrate specificity than other CoV proteases. Highlighted is the newly emerging putative NSP7+8 (2:1) trimer (orange) in the most intense peak series. Other NSP species detected at lower intensities encompass NSP7 (yellow), NSP8 dimer (green) and NSP7+8(1:1) heterodimer (grass-green). Complex stoichiometry confirmed via tandem MS (Figure 42 and Figure 43). In **Table S 1: Molecular weights (MWs)**

In native mass spectra of complex formation, the most intense peak series was assigned to NSP7+8(2:1) trimer (40.46 kDa for TGEV and 40.49 kDa for FIP) (Figure 41). Additionally, a

NSP7+8(1:1) dimer was detected. Based on the MWs of NSP7 and NSP8, the stoichiometry could be predicted (e.g. TGEV NSP7 2x9.5 kDa + NSP8 1x21.47 kDa = 40.47 kDa; FIP NSP7 2x9.5 kDa + NSP8 1x21.48 kDa = 40.48 kDa). For further evidence tandem MS was performed.

As CID precursors, the +13 main peaks (3112 *m/z* TGEV and 3105 *m/z* for FIP) were selected. The obtained product ion spectra gave further insight into the arrangement of the complexes (Figure 42 A and Figure 43 A). Other ions of these peak series were also subjected to CID (Figure S 7).

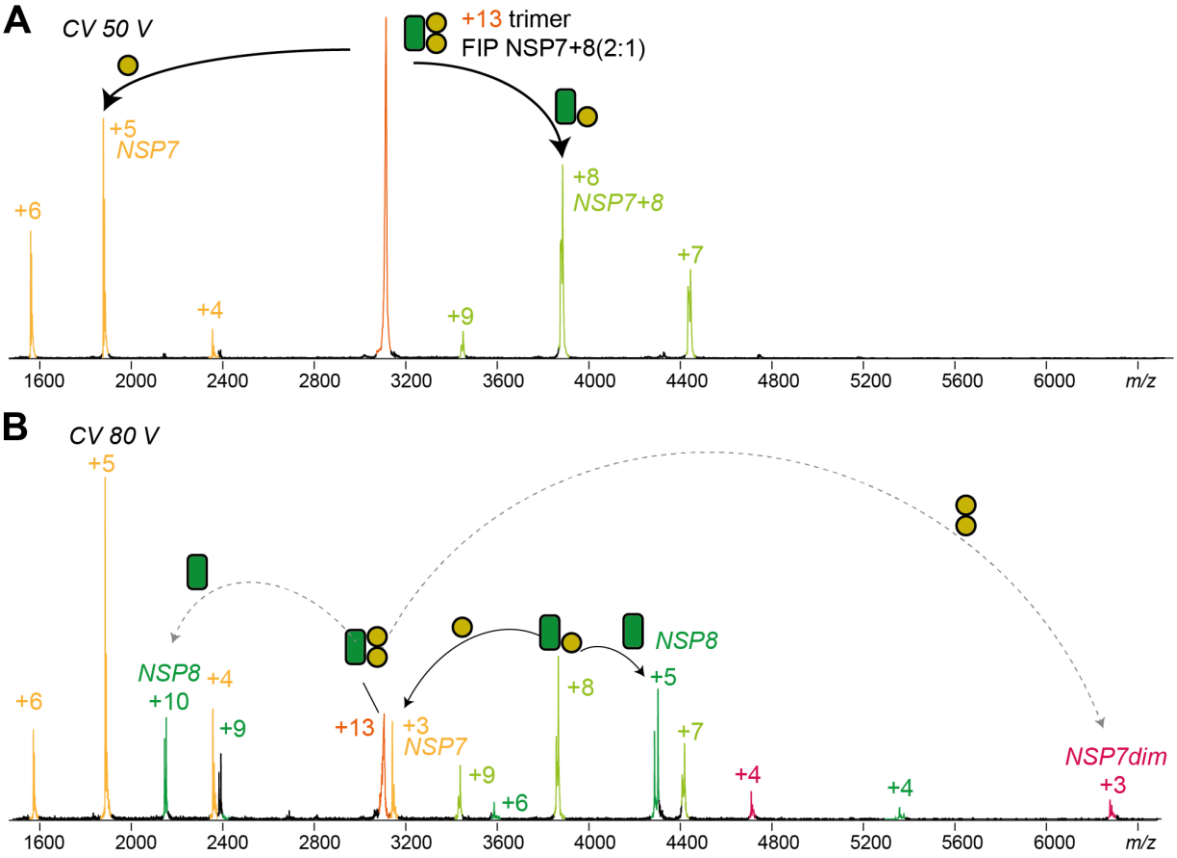


Figure 42: Collision-induced dissociation of FIP NSP7+8 trimer. Tandem MS of the +13 molecular precursor ion confirms trimers. Dissociation triggered in Q-TOF tandem MS with increasing CV at 14 µbar argon. Arrows depict matching dissociation products. **(A)** Product ion spectrum at CV 50 V showing dissociation of NSP7+8(2:1) into NSP7 (yellow) and NSP7+8(1:1) (grass-green) remainder. **(B)** Product ion spectrum collisional energy elevated to CV 80 V showing two further dissociation processes; first (black arrows), 2nd gen. dissociation of the NSP7+8(1:1) remainder into NSP7 and NSP8 (green) and second (dotted arrows), dissociation of NSP7+8(2:1) into NSP8 and remainder NSP7(2) dimer (crimson). Ejection of NSP7(2) dimer demonstrates the NSP7:NSP7 interface within the trimer. Double-peaks of FIP NSP8 derive from bivalent mass distribution of 80 Da. Arrows depict matching dissociation products. Precursor completely dissociated at CV 100 V (Figure S 11).

Two alternative pathways of dissociation were observed. Initially, at lower collisional activation, both trimers dissociated into NSP7 and remainder ions NSP7+8(1:1) similarly (Figure 42 A and Figure 43 A). Then, upon elevated collisional energy the remainder ions NSP7+8(1:1) ejecting the second NSP7 as high charged ions and NSP8 as low charged ions (Figure 42 B and Figure 43 B). Therefore, the order of dissociation followed successive release of the two smaller NSP7 subunits (Figure 44).

At elevated collisional energy an alternative dissociation pathway highlighted further similarities (Figure 42 B and Figure 43 B). In both trimers a fraction partitioned into NSP8 and NSP7(2) dimer. The NSP7:NSP7 interaction reveals clear differences of FIP and TGEV compared to SARS.

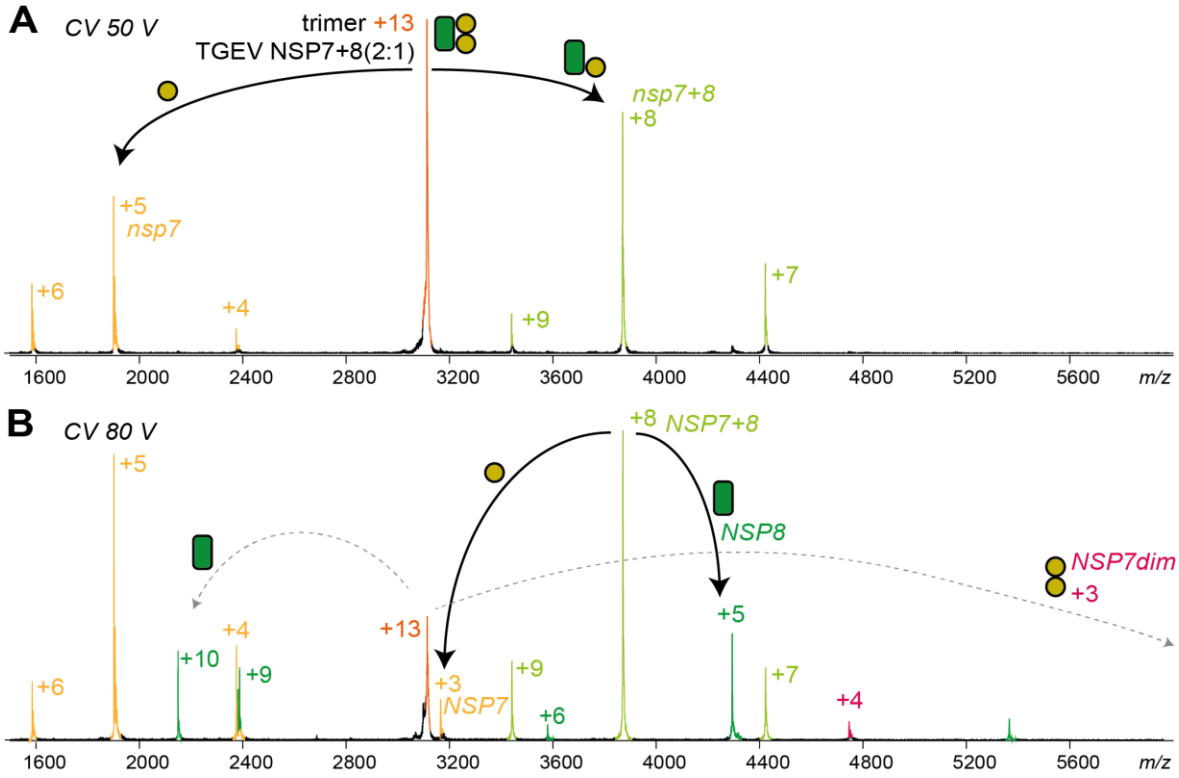


Figure 43: Collision-induced dissociation of TGEV NSP7+8 trimer. Tandem MS of the +13 molecular precursor ion confirms a non-covalent complex. Dissociation triggered in Q-TOF tandem MS with increasing CV at 14 µbar argon. Arrows depict matching dissociation products. The dissociation experiments confirm that native MS assigned NSP7+8(2:1) species is a non-covalent complex. **(A)** Product ion spectrum at CV 50 V showing dissociation into NSP7 (yellow) and NSP7+8(1:1) (grass-green). **(B)** Product ion spectrum at collisional energy elevated to CV 80 V showing, a similar dissociation pattern as FIP trimer; first (black arrows), follow-up dissociation of the NSP7+8(1:1) remainder into NSP7 and NSP8 (green), and second (dotted arrows) dissociation into NSP8 and remainder NSP7(2) dimer (crimson). Arrows depict matching dissociation products. Precursor completely dissociated at CV 100 V. The NSP7(2) dimer out of range here (Figure S 11).

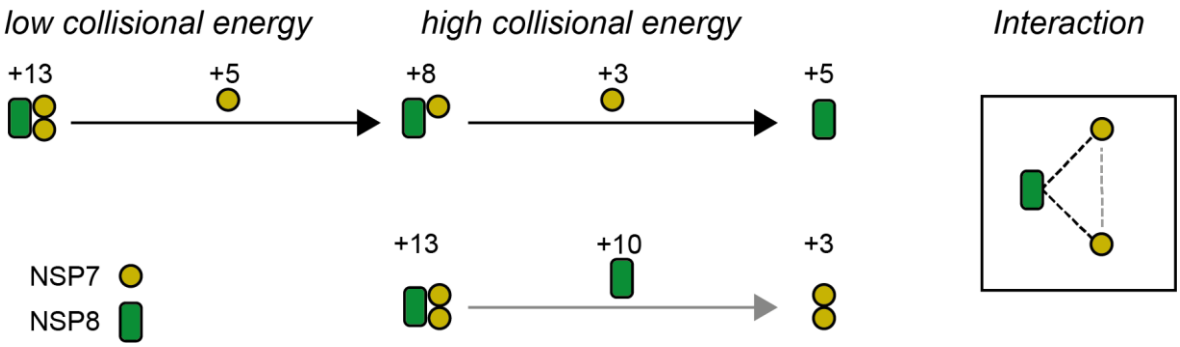


Figure 44: Gas-phase dissociation pathways of FIP and TGEV NSP7+8(2:1). Two alternative dissociation pathways from CID spectra of TGEV and FIP trimer and interaction of subunits within the trimer as concluded. Schematic of observed dissociation pathway of +13 NSP7+8(2:1) molecular ion. Symbols illustrate molecular ions from different mass species found as dissociating ions in CID. Black arrows indicate for the ejected ions. Charge states labelled taken from dissociating ions found in CID FIP (+13) and when combined match the charge state of the precursor ion. **First pathway (top):** At lower and elevated collisional energy consecutive loss of two NSP7 subunits. **Second pathway (bottom):** At elevated collisional energy, dissociation of NSP8 leaving the NSP7(2) dimer. **Interaction map:** Subunit interactions within the trimeric complex as suggested by the CID results. Continuous drawn arrows indicate for observed products.

Altogether, native MS showed NSP7+8 preferably forming a trimeric complex assembled of three individual subunits and within this complex both NSP7 interacting with NSP8 and with each other. This description suited a trimer of FIP NSP7+8, which is based on X-Ray crystallography and has been reported elsewhere [84]. Within this structure, NSP8 is bound to two NSP7 (NSP7I and NSP7II) molecules at different positions on its N-terminal elongated domain. The tighter binding NSP7I is grasping around the NSP8 shaft with several hydrophobic patches and at least two hydrogen bonds, while the weaker binding NSP7II forms only hydrophobic patches and is involved in positioning of the NSP8 N-terminal domain. Here, upon CID of the complex, successive dissociation of the two NSP7 domains was observed. From the trimeric FIP structure, it is most likely that at low collisional energy NSP7II is lost while NSPI is resistant due to its polar contacts granting gas phase stability. However, the interface of NSP7I:NSP7II consists of hydrophobic interactions and three hydrogen bonds, therefore it was not surprising to find the NSP7(2) dimer as dissociation product, too.

In contrast to the FIP complex, TGEV NSP7+8 has not yet been characterized before. The NSP7-8 protein had high sequence identity (93.9 %) with FIP, and stoichiometry and gas-phase dissociation were similar. To illustrate the similarities, a model was generated with SWISS MODEL and the highest prediction probability was reached for homologues of the FIP NSP7+8(2:1) (Figure 45). Due to the here provided experimental validation through native MS and CID, this model was considered as convenient for further research.

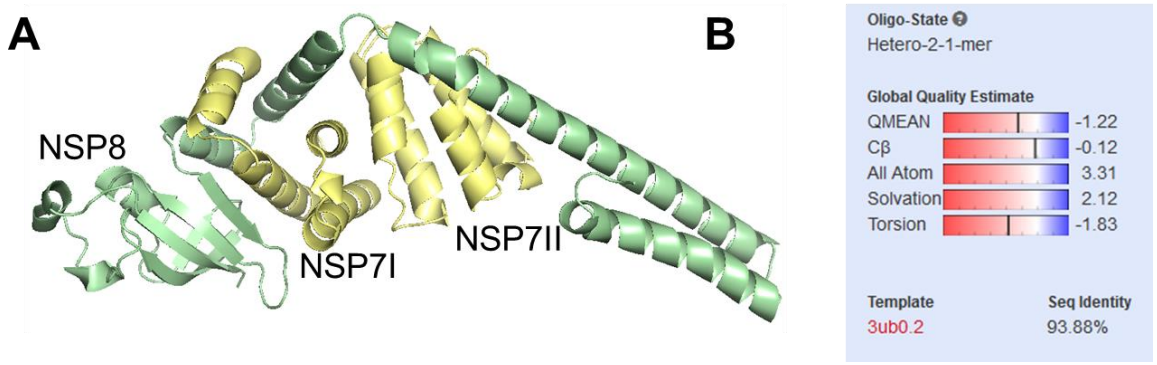


Figure 45: Structural homology model of TGEV NSP7+8(2:1) as confirmed by tandem MS. (A) Predicted molecular structure of TGEV NSP7 (yellow) and NSP8 (green) in context of the NSP7+8(2:1) trimer. **(B)** Parameters for homology modeling with SWISS-MODEL workspace. Input data were the TGEV amino acid sequence of NSP7 and NSP8. The molecular structure with the highest QSQE factor was suggested based on FIP NSP7+8(2:1) trimer (PDB 3UB0).

Altogether, the experiments showed that complex formation could be robustly reenacted. Consequently, the analysis was extended to NSP7+8 of PEDV and 229e, two yet unstudied homologues with much less sequence similarity.

4.3.3 Complexes of NSP7+8 from PEDV and 229e

To complete this project, NSP7+8 of PEDV and 229e were tested. Their molecular structure, complex stoichiometry and subunit arrangement had been unknown. Of the homologues analyzed in this thesis, PEDV and 229e were found to be part of one phylogenetic branch (Figure S 16). However, their NSP7-8 had only 70.9 % sequence identity, which is much less than 93.9 % for TGEV and FIP. This raised the question if PEDV and 229e also behaved similar regarding their complex parameters. Moreover, sequence identity of PEDV and 229e NSP7+8 was higher to TGEV and FIP than to SARS (Sequence identity matrix: Figure S 4). Whether their complex stoichiometry groups with one of the other viral species was of interest. To investigate complex formation, the native MS approach was used as mentioned before (Section 4.3.1). In the following paragraphs, I report results of, first optimization of experimental conditions and complex formation in native MS of both protein homologues, then CID-MS of different PEDV and 229e complexes and at last, one analysis of an unspecific cleavage product that gave insight about binding sites.

In order to start working on PEDV and 229e proteins optimization of two experimental parameters was necessary, namely reducing agent and ionic strength. Initially, in native MS both uncleaved NSP7-8-SGSG-His proteins existed as monomers (32.0 kDa; 32.0 kDa) and dimers (64.1 kDa; 64.0 kDa). Their ratio did not significantly change with increasing concentration. To test for non-covalent interactions, CID of the dimers was performed (Figure 46). At collisional energies sufficient to deplete similar sized precursor ions, only sparse

dissociation of the PEDV precursor ion occurred. Instead, the protein underwent fragmentation. This result suggests a high dimer stability, which could have occurred due to disulfide bridges caused by storage conditions during protein shipping. To reduce possible disulfides, the reducing agent concentration was increased to 5 mM DTT before analysis. In general, 229e proteins had been cumbersome to work with being prone to aggregation. They stabilized at increased salt conditions, which made complex analysis possible in the first place. Therefore, the NSPs were electrosprayed from slightly deviating conditions, having salt concentration increased to 500 mM AmAc instead of 250 mM as used in other experiments in this thesis.

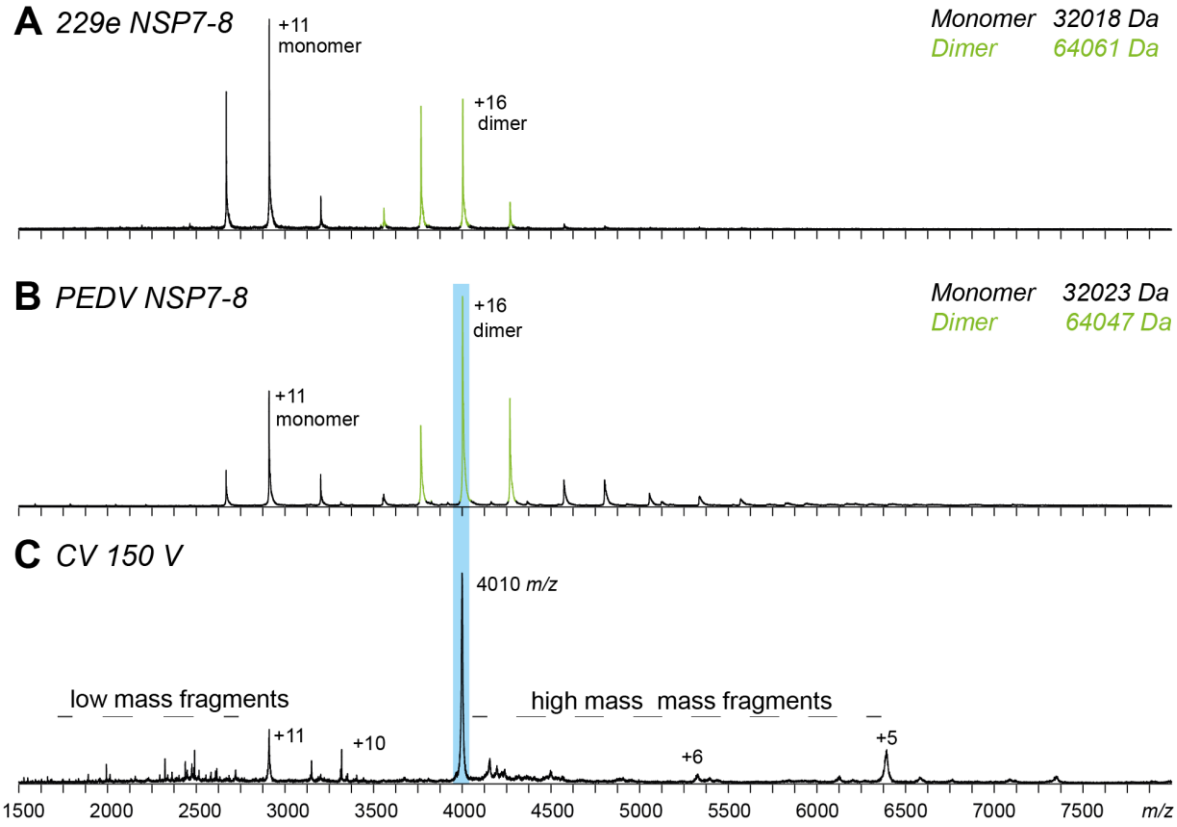


Figure 46: Highly stable PEDV and 229e NSP7-8 dimer. Native mass spectra showing partial dimerization of NSP7-8-SGSG-His of **(A)** 229e and **(B)** PEDV. Monomer-dimer (black-green) equilibrium did not change upon lowering the concentration. **(C)** Precursor and product ion spectrum of PEDV NSP7-8 dimer (+16, 4010 *m*/*z*) at CV 160 V showing dissociation of non-covalent subunits as well as onset of fragmentation. Fragments most likely derive from dissociating monomers. Indicted are precursor ion selection for CID (blue bar) and product fragments (horizontal dashed lines). Proteins electrosprayed at 10 μM in 500 mM AmAc, 1 mM DTT at pH 8.

In the main experiment, polyproteins NSP7-8-SGSG-His were cleaved by M_{pro} tailoring the proteins to contain authentic ends (Protein constructs listed in Table 3). Native mass spectra of PEDV and 229e NSP7+8 appeared virtually similar regarding the assignment complex species and their relative intensities (Figure 47). The most abundant mass species were NSP7+8(2:2) tetramer (61.9 kDa PEDV; 61.9 kDa 229e). Additionally found were NSP7+8(1:1) dimers (30.9 kDa PEDV; 30.9 kDa 229e) and to a much lesser degree NSP7+8(2:1) trimers (40.1 kDa PEDV; 40.2 kDa 229e). By all and large, these findings resembled complex formation of SARS which also formed preferably hetero-tetramer. The tetramer stoichiometry was predicted based on the MW of NSP7 and NSP8 (2x9.2 kDa + 2x21.7 kDa = 61.8 kDa PEDV; 2x9.3 kDa + 2x 21.6 kDa = 61.8 kDa 229e). Additionally, tandem MS was required to obtain more evidence and to shed light on subunit arrangement.

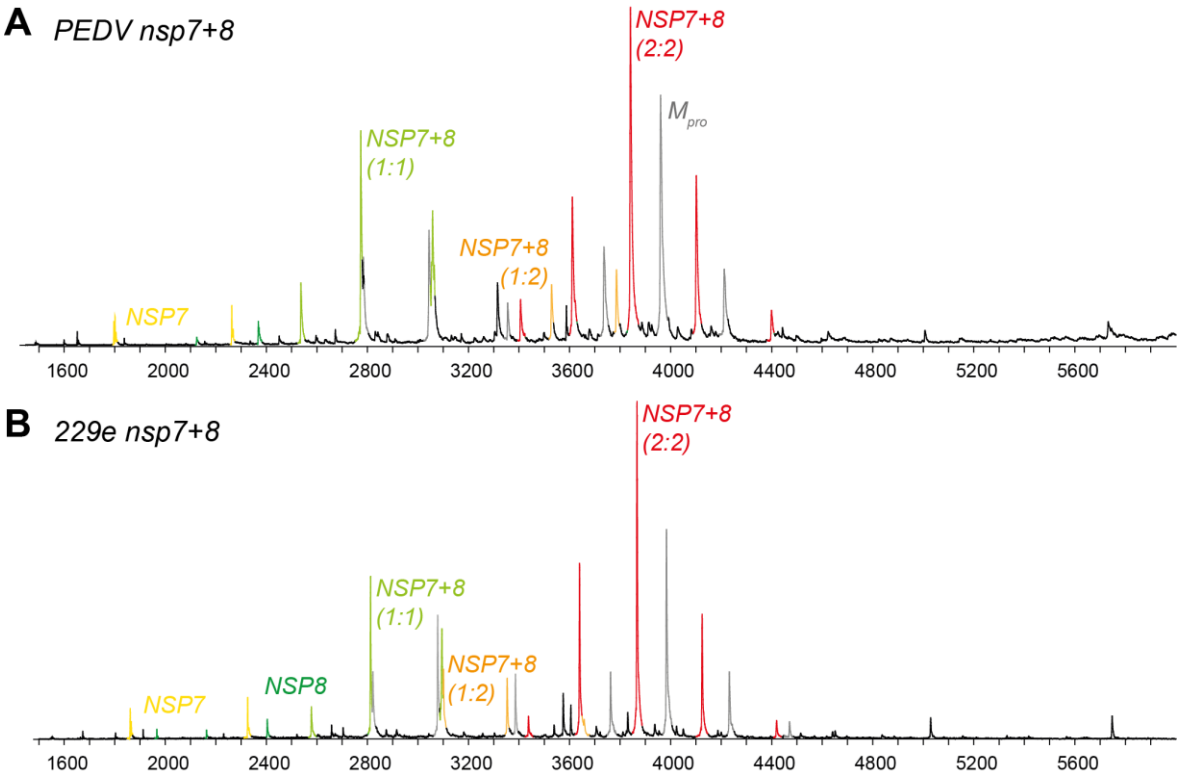


Figure 47: 229e and PEDV NSP7+8 assembles into (2:2) hetero-tetramer. Native mass spectra showing complexes of NSP7+8 of **(A)** PEDV and **(B)** 229e. For cleavage SARS M_{pro} was separately incubated with 20 µM PEDV and 20 µM 229e NSP7-8-SGSG-His (ratio 1:10) at 4°C in 500 mM AmAc, 2 mM DTT, pH 8 overnight. Complex formation is not influenced differently by the non-natural protease, which has virtually similar substrate specificity than other CoV proteases. PEDV and 229e exhibit similar complex formation. Highlighted are complexes of NSP7+8(2:2) hetero-tetramer (red). Further mass species were assigned at medium intensity, hetero-dimer (grass-green), NSP7+8(2:1) trimer (orange), and SARS M_{pro} (grey), as well as at lower intensity, NSP7 (yellow) and NSP8 (green). Overlapping species between 2800 *m/z* and 3600 *m/z* impeded mass assignment of more than three consecutive peaks of NSP7+8(2:1). Stoichiometry of complexes confirmed via tandem MS (Figure 42 and Figure 43). (**Table S 1: Molecular weights (MWs) of NSPs, their cleavage products and complexes.**)

To learn about subunit stoichiometry of PEDV and 229e complexes, tandem MS was performed. The observed dissociation pathways between proteins of both viral species partially differed, but similarities were predominant and therefore are reported in conjunction. As precursor ions for CID, hetero-tetramers were selected at different charge states for PEDV (+16, 3866 *m*/*z*; +17, 3638 *m*/*z*) and 229e (229e +15, 4124 *m*/*z*; +16, 3867 *m*/*z*; +17, 3640 *m*/*z*) (Figure 48 and Figure S 7).

Three pathways of dissociation were observed (Figure 49). First, the precursors dissociated into high charged NSP8 and low charged NSP7+8(2:1). Their peaks increased with the collisional energy until the precursor ion was depleted. In fact, the product peaks became so predominant that virtually no other peaks were visible in relation. NSP8 with high charge states (+11 and +12) reflected both, unfolding during CID and the capacity of charge withdrawal due to the elongated nature of this protein. Dissociation of exactly one NSP8, while the other one remained bound in NSP7+8(2:1), suggested a weak gas phase interaction and possibly even, different NSP8 binding modes within the complex. It is more likely that the remainder complex NSP7+8(2:1), which had been stripped of charges, had a too low energy and therefore, was unable to further dissociate resulting in 2nd generation (gen.) products.

A second pathway of dissociation was observed at elevated collisional energy when lower charge states of NSP8 (+7) were ejected (Figure 48 A and B). This led to less withdrawal of charges from the remainder NSP7+8(2:1) and set the stage for 2nd gen. products. Thereby, ejection of the second NSP8 (+5) molecule generated a NSP7(2) (+5) dimer, indicating a central location within the complex. Surprisingly, the stability of specifically 229e NSP7(2) was higher than that of PEDV NSP7(2) at high collisional activation. Since the gas phase stability is not directly comparable to that in solution, differences in dissociation of PEDV and 229e must not stand for a different arrangement of subunits. However, dissociation of the NSP7(2) dimer shows a clear difference to the SARS tetramer but suggests similarity to TGEV and FIP trimer, which exhibited a similar dissociation pattern (Section 4.3.2).

A third pathway of dissociation was found in both species, but more predominantly with the PEDV tetramer. A second charge envelope of NSP8 arose (+4 to +7) with even less charges, allowing the remainder to eject a monomeric NSP7 (+3 to +4) and thus, resulted in a NSP7+8(1:1) (+6 to +8) 2nd gen. product ion. This ion was evidence of a tetramer assembled of four individual subunits and that in fact, a relatively stable interaction of NSP7+8(1:1) existed. For the 229e tetramer the products of this pathway remained relatively low intense. One reason could be that NSP7 monomer signaled at *m*/*z* values already populated by intense signals of the NSP7(2) dimer (NSP7: +3, 4651 *m*/*z*; 3101 *m*/*z*; NSP7(2): +6, 4651 *m*/*z*; +8, 3151 *m*/*z*) and therefore, the monomer dissociated in fact more than the assigned peaks suggest. However, besides of the depicted differences in the third pathways, CID of the 229e tetramer showed more similarity to PEDV than to other tetramers tested.

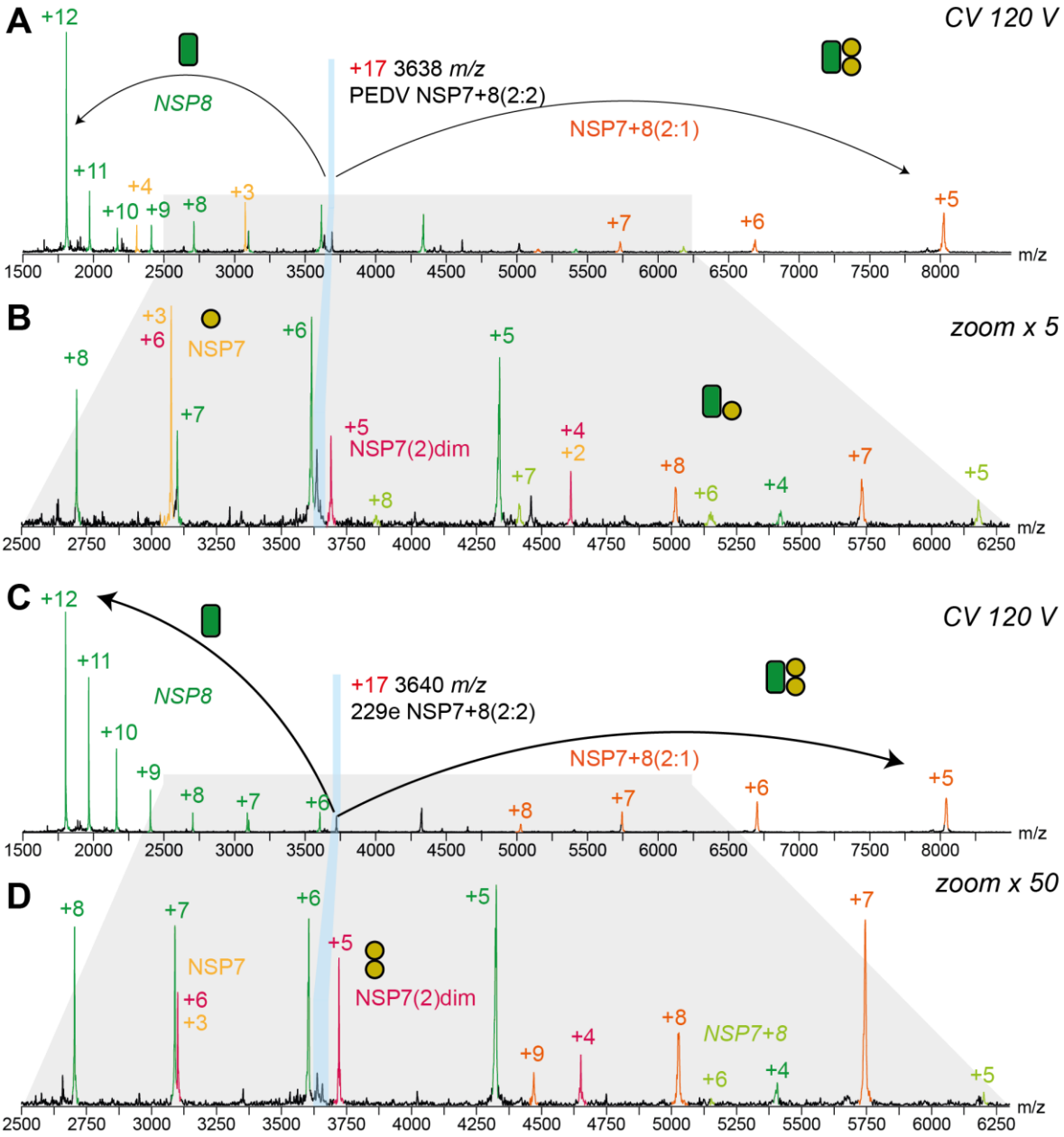


Figure 48: Collision-induced dissociation of PEDV and 229e hetero-tetramer. Dissociation triggered in Q-TOF tandem MS with increasing CV at 14 µbar argon. Arrows depict matching dissociation products. **(A)** Product ion spectrum of NSP7+8(2:2) (+17, 3638 *m/z*) from PEDV at CV 120 V and **(B)** 5x magnification (below, 2500 – 6300 *m/z*). **(C)** Product ion spectrum of NSP7+8(2:2) (+17, 3640 *m/z*) from 229e at CV 120 V and **(D)** 5x magnification (below, 2500 – 6300 *m/z*). At CV 120 V the precursor ions almost completely dissociate. Arrows indicate for charge states matching the precursor ion. Both complexes show first dissociation of NSP8 (green, +9 to +12) and NSP7+8(2:1) (orange, +5 to +6), as high and low charged species, respectively. Notably, the complexes eject first the higher mass ion NSP8, indicating peripheral positioning within the complex. Further, product ions reveal protein interaction within the complex such as NSP7(2) dimer (crimson) or NSP7+8(1:1) (grass-green). The difference is that CID of dissociation of tetramer into NSP7+8(1:1) and NSP7 monomer was much stronger in PEDV.

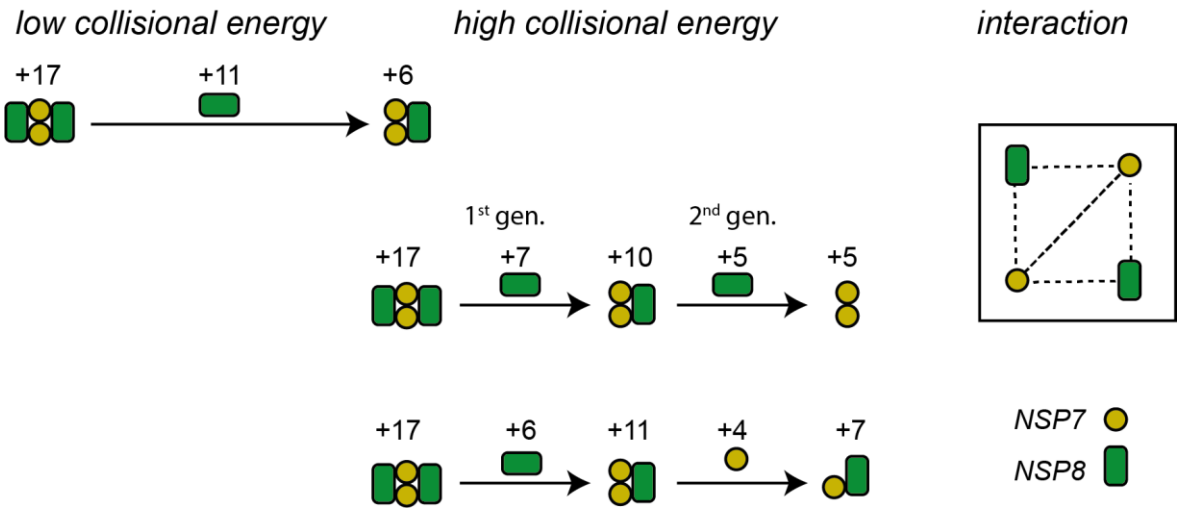


Figure 49: Gas-phase dissociation pathways of NSP7+8(2:2) from PEDV and 229e. Three dissociation pathways from CID spectra of PEDV and 229e tetramers and interaction of subunits within the tetramer as concluded. Symbols illustrate molecular ions from different product ions in CID. Arrows indicate for the ejected ions. Charge states are labelled and were taken from products in CID of PEDV tetramer (+17). **First pathway (top):** Tetramer ejects NSP8 due to unfolding leading to extensive charge partitioning and extremely low charged NSP7:8(2:1). **Second pathway (middle):** Tetramer subsequently ejects two NSP8 subunits leaving NSP7(2) dimer as 2nd gen. remainder ion. **Third pathway (bottom):** Tetramer ejects one NSP8 followed by one NSP7 leaving NSP7+8(1:1) as 2nd gen. remainder ion. **Interaction map:** The reminder ions indicate for specific subunit interaction within the complex. CID-MS demonstrated NSP7(2) and NSP7+8(1:1) interaction.

Altogether, stoichiometry of the complexes group with the arrangement of SARS NSP7+8(2:2). However, in CID, particularly the product ion NSP8(2) dimer, which was typically for the SARS tetramer scaffold, was not observed for PEDV and 229e. Instead a NSP7(2) dimer was detected resembling FIP and TGEV. Therefore, tetramer arrangement is indicated differently from SARS and rather a FIP/TGEV NSP7+8(2:1) arrangement with an additional NSP8. One likely possibility could be an arrangement of a centered NSP7(2) dimer flanked by two peripheral NSP8. Further proof for this arrangement was obtained by tandem MS of PEDV and 229e trimers and dimers.

To learn more about the complexes, tandem MS was performed of 229e NSP7+8(1:1) and NSP7+8(2:1). In one particular example, two overlapping precursor peaks were selected (+10, 3093 *m/z*; +13, 3095 *m/z*) (Figure 50 A), and the dissociation pathway of both, dimer and trimer, became clear.

Initially and as expected at low collisional energy, the dimer dissociated into NSP7 and NSP8. Meanwhile, the trimer dissociating into NSP7 and NSP7+8(1:1) (Figure 50 B). Then at elevated collisional energy, the trimer dissociated in an alternative pathway into NSP8 and NSP7(2) dimer (Figure 50 C). These results showed that the NSP7 monomer was part of the 229e trimer and further, that within this complex the NSP7(2) dimer interacted with a high gas phase stability.

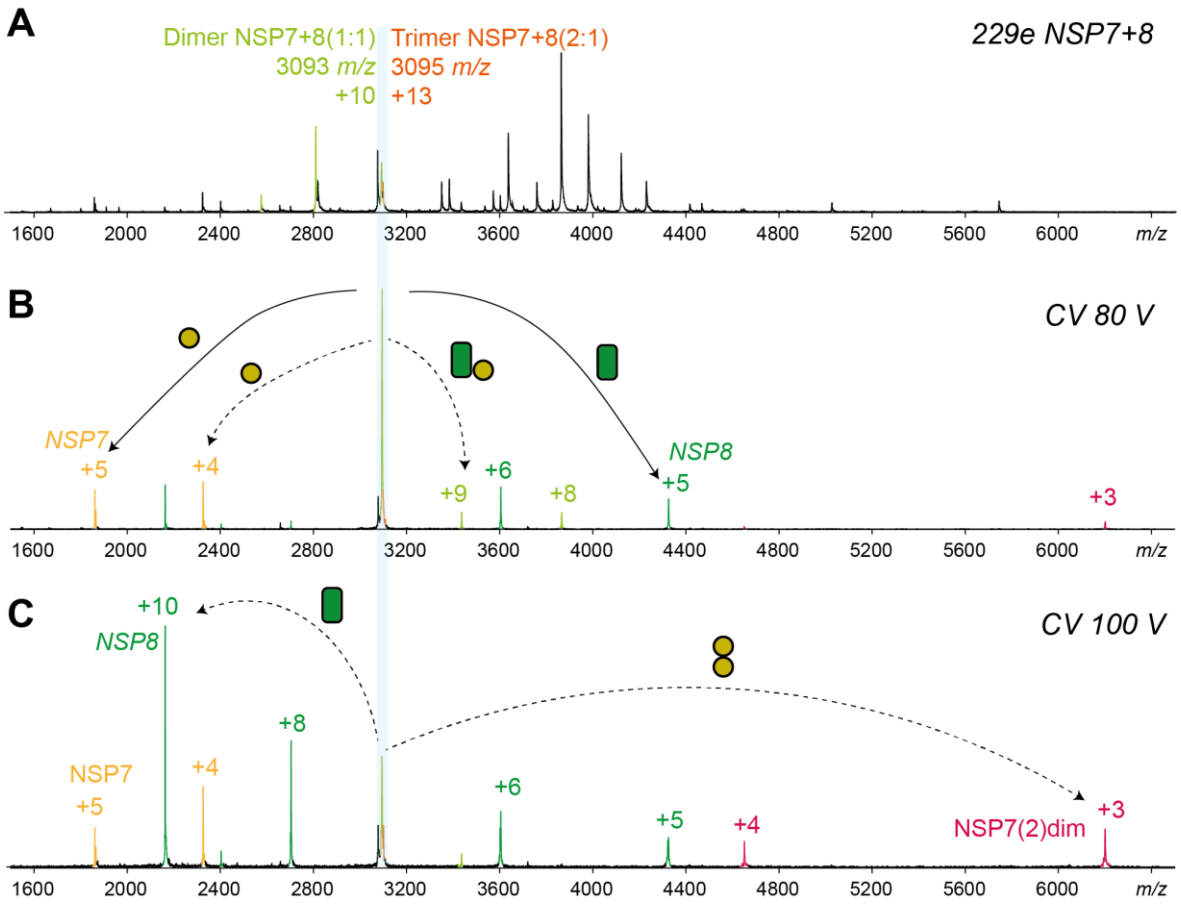


Figure 50: Collision-induced dissociation of 229e NSP7+8 dimer and trimer. (A) Native MS overview spectra of 229e NSP7+8 complex formation lead to overlapping peak signals from 3000 *m/z* and 3200 *m/z* later assigned to NSP7+8(1:1) dimer (grass-green) and NSP7+8(2:1) (orange). Precursor ions of both species encompassing 3093 *m/z* (blue bar) were quadrupole selected and subjected to collision cell with 14 µbar argon. **(B)** Product ion spectrum at CV 80 V shows that the dimer dissociates (black arrow), as expected for NSP7+8(1:1), into low charged NSP7 (yellow) and high charged NSP8 (green). Additionally, the trimer dissociates (dashed arrow) into NSP7 and NSP7+8(1:1) (grass-green). **(C)** Product ion spectrum at CV 100 V shows that the trimer alternatively dissociates (dashed arrow) into NSP8 and NSP7 dimer (crimson), as high and low charged ion species. Arrows indicate matching dissociation products.

Comparing these results to the dissociation of the 229e tetramer, where NSP7(2) was virtually the only follow-up dissociation product detected, it appears that the limited dissociation was in fact a result of charge withdrawal. Further, the gas-phase dissociation of the trimeric 229e is very similar to trimeric FIP and TGEV (Figure 42) and therefore, the results support the picture that the 229e tetramer arrangement resembles FIP/TGEV NSP7+8(2:1) with an additional NSP8.

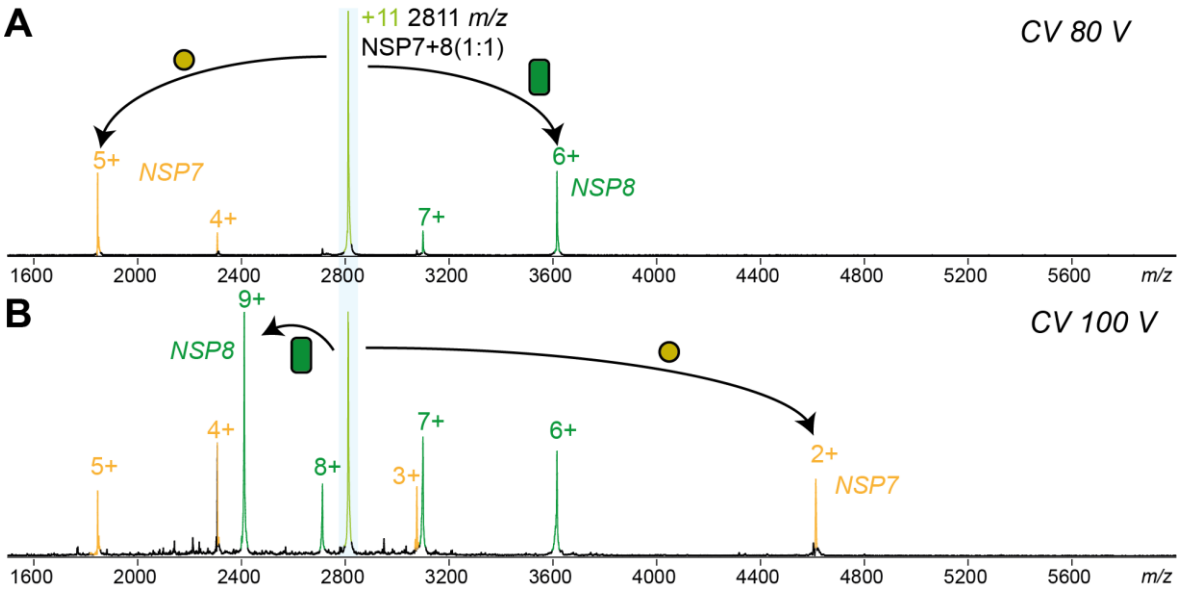


Figure 51: Switch of charge partitioning upon CID of PEDV NSP7+8 dimer. Product ion spectra of NSP7+8(1:1) (+11, 2811 *m/z*) from PEDV. Molecular ions encompassing 2811 *m/z* (blue box) were quadrupole selected and subjected to the collision cell at 14 µbar argon at increasing acceleration energy. The NSP7+8(1:1) dimer (grass-green) undergoes two pathways of asymmetric charge partitioning. **(A)** At lower collisional energy (CV 80 V), dissociation of NSP7 (yellow) and NSP8 (green), as high and low-charged ions respectively. **(B)** At elevated collisional energy (CV 100 V), charge partitioning between NSP7 and NSP8 is reversed. The latter pathway might result from disruption of NSP7+8 binding unfolding of NSP7 and hence, withdrawal of the complexes' charges by the naturally elongated NSP8 fold. Arrows indicate for matching dissociation products.

Continuing with CID, dissociation of PEDV NSP7+8(1:1) hetero-dimer (+11, 2811 *m/z*) was tested. The results classify as a native MS oddity teaching about unfolding and subunit ejection from a two-modal ejection mechanism. First, at lower collisional energy, the dimer dissociated accordingly into NSP7 (+5) and NSP8 (+6) (Figure 51 A). This demonstrated the asymmetric charge partitioning as expected, because the small subunit NSP7 unfolds first and withdrawals charges from the complex. Then at elevated energy, an unexpected mechanism was observed, when charge partitioning swapped, and the dimer dissociated accordingly into NSP7 (+2) and NSP8 (+9) (Figure 51 B). It appears that when energy is deposited from higher energy collisions dimer interactions disrupt before NSP7 unfolds. Thereby, with NSP7 unable to withdraw charges, NSP8 remains highly charged.

The concluded mechanism suits well to the elongated structure of the NSP8 homologues, which allow for a small binding interface in relation to their molecular weight. The results supported the molecules elongation and thereby the structural similarity between NSP8 of different coronavirus species. Furthermore, disruption before unfolding in the dimer also reflected the high charge products of NSP8 as observed in the tetramer. Altogether, results from CID of the dimer and trimer concordantly added to the conclusions drawn from CID of the tetramer.

4.3.4 *Complex formation of NSP7 with a truncated NSP8 from PEDV*

One specific analysis of a truncated PEDV protein gave further insight into binding sites (Figure 52). This truncation was based most likely on unknown proteolytic activity introduced during sample preparation.

Initially, the experiment encompassed M_{pro} cleavage of NSP7-8-NNE-His protein as a precursor for complex formation of NSP7+8. Native mass spectra of this sample showed a high heterogeneity of NSP products and complexes (Figure 53).

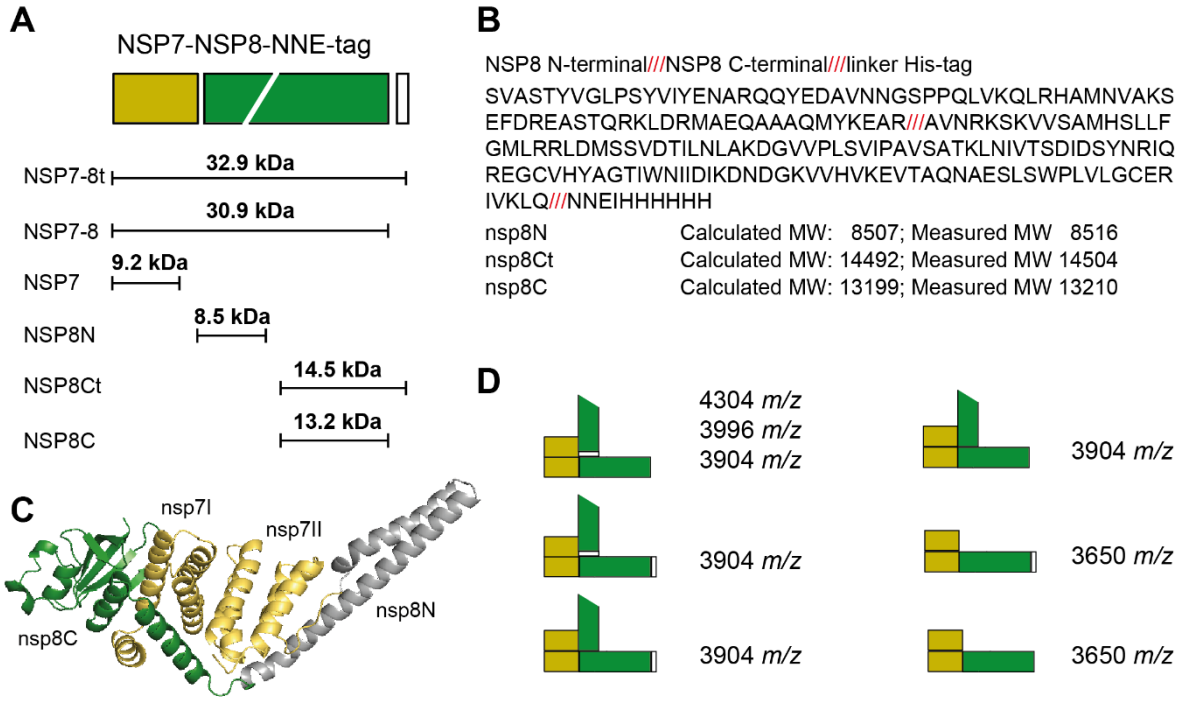


Figure 52: Complex formation of a truncated NSP7-8 from PEDV. (A) Schematic and terminology of the mass species detected upon native MS of M_{pro} cleaved PEDV NSP7+8-NNE-His. **(B)** Comparing theoretical and determined mass species allowed for location of the truncation within the amino acid sequence. **(C)** A homology model of a possible PEDV trimer (2:1) with truncation site mapped by color. Tetrameric stoichiometry is not represented in this model. The structure suggests tight NSP8C (green) binding to NSP7I. **(D)** Symbols of complexes found in tandem MS with precursor ions are indicated with NSP7 (yellow), His tag (white) as well as NSP8 and NSP8C (green).

Next to M_{pro}, merely two other mass species were assigned unequivocally, NSP7 monomer and N-terminally truncated NSP8, termed NSP8N (8.5 kDa). Later in tandem MS, the NSP8 C-terminal counterparts were found as well, termed NSP8C (13.2 kDa) and NSP8Ct (14.5 kDa), with and without His-tag, respectively. The mass of the products revealed the site of truncation after an arginin residue (MYKEAR///) and a predicted model of PEDV trimer located this site in the N-terminal loop domain of NSP8 (Figure 52 A-C, Figure S 12).

Striking, NSP8C/NSP8Ct was found in CID of complexes, while NSP8N was only found in overview spectra (Figure 53 A-E). These findings highlight that on the one hand, the NSP8 C-

terminal binding site is critical for interaction with NSP7 and on the other hand, the insufficiency of NSP8N for binding.

Furthermore, tandem MS was performed on four precursor ions encompassing unassigned peaks (3650 *m*/*z*, 3904 *m*/*z*, 3996 *m*/*z*, 4303 *m*/*z*). The mass of dissociating ions increased with its precursor ion`s *m*/*z* and showed evidence for at least six different complexes (Figure 52 D). In all CID spectra, NSP8C/NSP8Ct and NSP7 were found as products which demonstrates their presence in trimeric and tetrameric complexes. Unfortunately, lack of 2nd gen. products hindered identification of uncleaved NSP7-8 or NSP7+8(1:1) dimer complexes. Some of the CID spectra showed dissociation of full-length NSP8 suggesting that the assigned mass species were, in fact, complexes.

Interestingly, NSP8c/NSP8ct was frequently ejected in CID (Figure 52 B-E). This process resembled NSP8 dissociation of the PEDV/229e tetramer but fewer charges were withdrawn due to truncation of the elongated NSP8 domain. Furthermore, it was found that the complexes contain two NSP7 molecules. This second NSP7 appeared to be a prerequisite for protein-protein interaction with NSP8c/NSP8ct. Therefore, it was concluded that one way of forming a NSP7+8(2:2) PEDV hetero-tetramer was first binding of NSP7+8(1:1) dimer to a second NSP7, which only then allowed binding of a second NSP8.

Moreover, Li et al. reported a 'resectioned' NSP8 [87] but could only roughly locate the site to the loop region of NSP8. In the crystal structures of SARS hexa-decamer, only half of NSP8 have such a loop and the other half have a straight α-helix. They speculated that only NSP8 with loop gets 'resectioned' and that it would then still be able to bind with the full-length NSP8 and thereby, tune enzymatic function of this complex. As shown in the results, the truncation was located here, and, in fact, NSP8c was able to bind in different stoichiometries to complexes with two NSP7 and one NSP8. However, the reason for this truncation, as well as its connection to any enzymatic switch remains elusive.

Altogether, this specific sample of truncated NSP8 highlighted the importance of the C-terminal domain of NSP8 for complex formation and further supported the structural arrangement of PEDV and 229e hetero-tetramer, as concluded above. In the end of this thesis, results of NSP7+8 complex formation are also discussed in context of the virological relevance (Chapter 5: Conclusion and outlook).

Native MS was used successfully to investigate NSP complex arrangement, and to endorse the drawn conclusions, additional methods of structural MS were consulted, as shown in the next chapters.

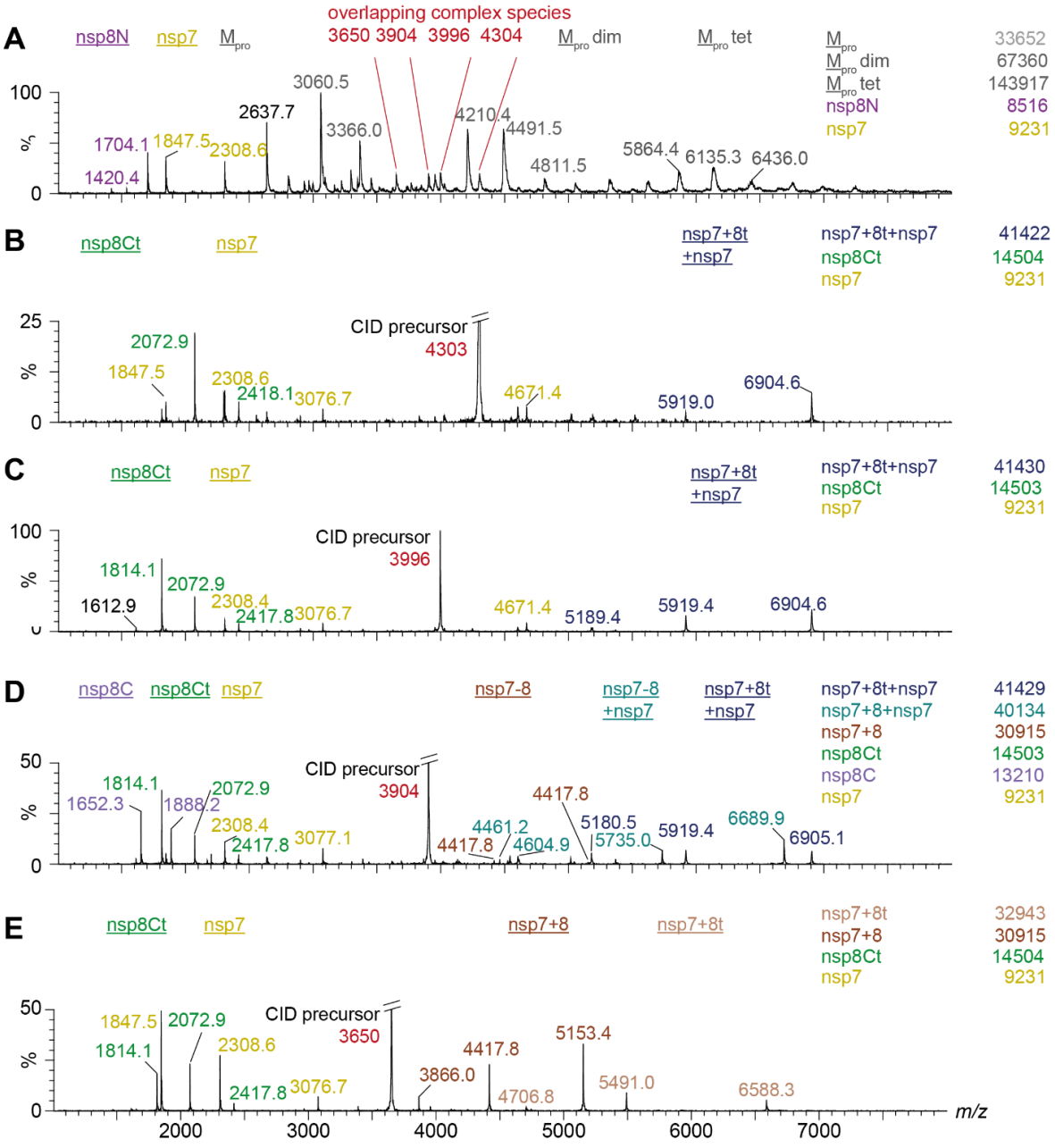


Figure 53: Native MS and collision-induced dissociation of a truncated PEDV NSP7-8. Native mass spectrum of 1 µM FIP M_{pro} incubated with PEDV 16 µM NSP7-8-NNE-His (ratio ~1:16) overnight and then diluted (1:1) in 250 mM AmAc, 1 mM DTT at pH8. High heterogeneity occurs due to products of multiple lengths including mass species with and without His-tag and an M_{pro} independent truncation in NSP8. NSP8N (purple), NSP7 (yellow) and M_{pro}(grey). Despite truncation multiple complexes formed. To identify complexes, unassigned signals were selected as precursor ions (red) for CID-MS; **(B)** 4303 *m/z* (CV 100 V), **(C)** 3996 *m/z* (CV 100 V), **(D)** 3904 *m/z* (CV 100 V) and **(E)** 3650 *m/z* (CV 80 V). NSP8N was not found in CID suggesting that it is not involved in protein interactions. MW (in Da) of mass species is given in legend. Based on the MWs of monomers the stoichiometry was reconstructed. For complex assignment at CID of 3650 *m/z*, NSP7 but not NSP8C was combined with NSP7-8t/NSP7-8 because of higher peak intensity of NSP7. Spectra normalized for their base peak, which is always the precursor ion. To simplify visualization, spectra were magnified and main peak cut accordingly, as indicated on the y-axis.

4.3.5 Sequence alignment of NSP7+8 interfaces

The analysis of NSP7+8 of different CoV species showed different stoichiometries and subunit arrangements, which can be sorted into three categories. Unlike SARS NSP7+8 tetramers (2:2), which are held together by a NSP8(2) dimer 'scaffold', the PEDV and 229e tetramers (2:2) had a central interaction of a NSP7(2) dimer. A third category of complexes was found in NSP7+8 of TGEV and FIP, which also had the central NSP7(2) dimer interaction but a trimeric (2:1) stoichiometry.

Molecular structures of SARS and FIP have been published that agree with the acquired results [86]. To search for conserved amino acids involved in the binding of these structures, a sequence alignment was conducted. As expected, the hydrophobic residues in the primary NSP7:NSP8 binding site for SARS and FIP are similar or conservatively replaced (hydrophobic SARS 19 of 20, FIP 17 of 21) between all viral species tested. The polar interactions indicated from SARS and FIP structures are conserved in one or the other species but do not follow any specific pattern (Figure 54).

Unique structural features of FIP were identified that could explain the predominantly trimeric stoichiometry of FIP and TGEV and the strong NSP7(2) dimer interaction shared by FIP, TGEV, PEDV and 229e. Most of the amino acids directly involved in these binding sites are similar or conservatively replaced in NSP7II:NSP8 (8 of 8 hydrophobic residues) and NSP7:NSP7 (7 of 9 hydrophobic residues, 3 of 5 polar interacting residues). Merely two amino acids fit in the target pattern of a residue that plays role in binding and is similar in all analyzed species except SARS; First, Ser18(NSP7) involved in a hydrogen bond in NSP7I:NSP7II and second Ile73(NSP7) involved in hydrophobic interactions of both, NSP7I:NSP8 and NSP7I:NSP7II. The latter (Ile73) is located in the loop region of NSP8, which appears to be a central point between all subunits of the trimer. Within this area, there are also two more residues (Glu71) and (Ser75), which fit into the target pattern, but were structurally not directly involved in binding. Remarkably, all amino acids which are directly involved in NSP7II:NSP8 binding in FIP were also conserved in SARS.

Consequently, residues of the SARS NSP8(2) dimer scaffold, as a unique feature of the tetramer, were analyzed regarding their conservation. Unfortunately, amino acids involved in binding were not specifically reported, and thus all residues near the interaction site in the structure, which point from one NSP8 towards the other NSP8, were chosen for comparison (7 of 10 hydrophobic residues, 0 of 3 other residues). It appears that the hydrophobic amino acids are conserved, and specifically one charged side chain (Lys165) near the interface is present in SARS, PEDV and 229e but not in TGEV and FIP. Since this side chain could not be assigned to a specific interaction, conclusions must be drawn carefully.

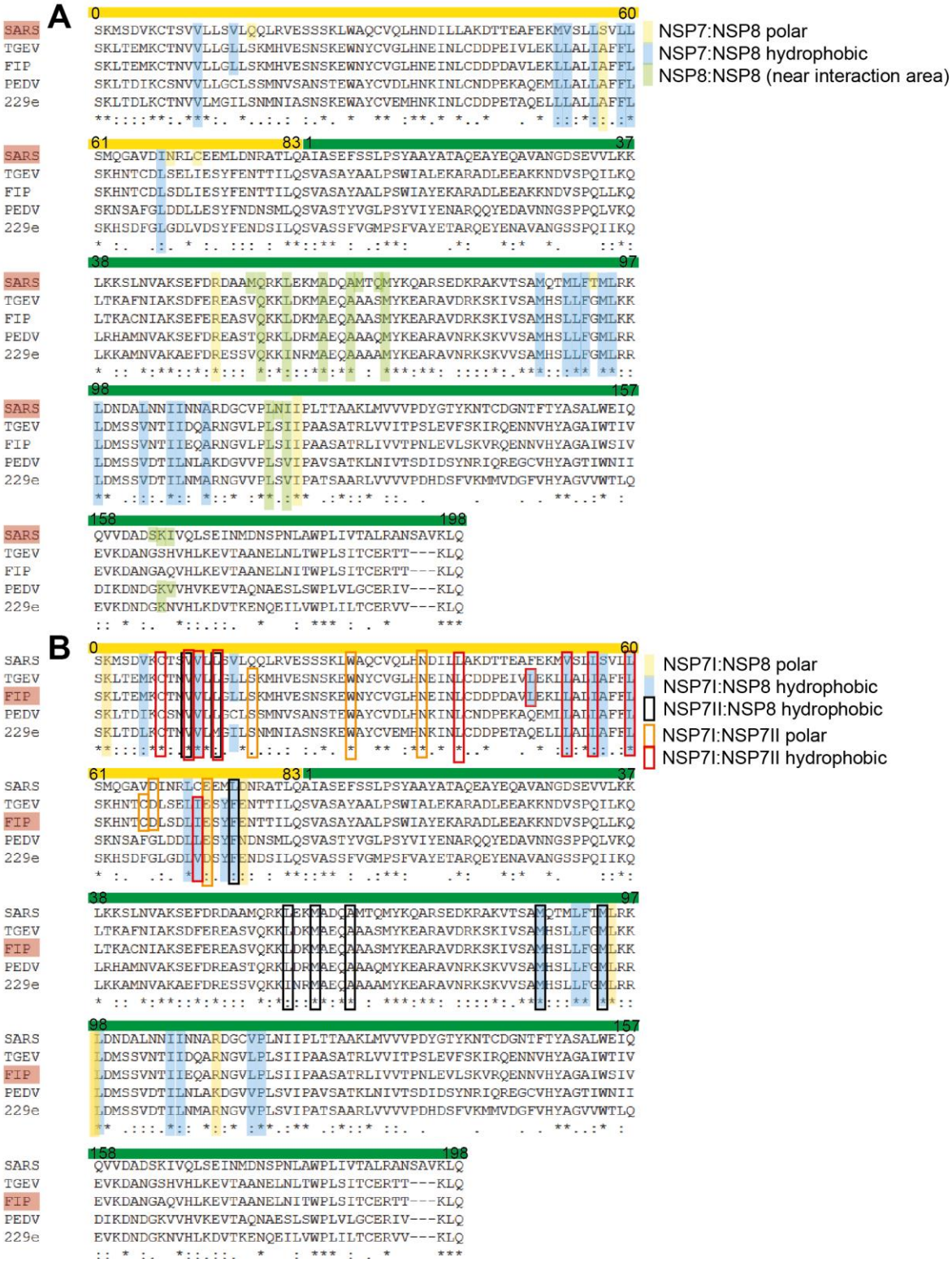


Figure 54: Analysis of conserved interacting amino acids in NSP7-8 proteins. Highlighted residues participate in binding interface in **(A)** SARS NSP7+8 according to [84] and **(B)** in the FIP NSP7+8 according to [86]. Different species interfaces are color coded according to the legend. Clustal Omega used for sequence analysis. MALDI-MS as a tool to investigate NSP processing and complex formation. (*)conserved residue, (:) conservative replacement, (.) semi-conservative replacement, ( ) not conserved. Table S 2: Amino acid sequences of pp1a/ab (region NSP4/5-NSP10/11/12).

4.4 MALDI-MS as a tool to investigate NSP processing and complex formation

During a COST fellowship (European cooperation in science & technology, Action 1403) at ETH Zürich, I had access to a high-mass MALDI (matrix-assisted laser desorption ionization) MS. With this opportunity in hand, I also checked NSP protein samples. This method allowed a different view because ionization disrupted non-covalent interactions and allowed observation solely of covalent cleavage products. Moreover, MALDI is based on a different principle and therefore may reveal different mass species than ESI. In this chapter, I present results of MALDI-MS analysis of 229e NSP7-10 polyprotein processing as well as FIP and 229e NSP7+8 complex formation.

To study processing order the polyprotein 229e NSP7-10-His was incubated with SARS M_{pro} for various times before the reaction was stopped by mixing with matrix (Non-natural protease for 229e substrate, but has an allover similar substrate specificity). MALDI was performed from dried sample-matrix spots. Even though mass spectra were not calibrated, all cleavage products of NSP7-10 could be clearly assigned as singly charged ions.

Comparing the relative intensities between time-points, three steps of polyprotein processing were clearly distinguished. First, the NSP7-10 and the protease peak were dominant. Additionally, rising signals for NSP10 and accordingly NSP7-9 intermediate indicated the highest cleavage activity at their junction. A second step of processing was signaled by depletion of NSP7-9 and rise of NSP8-9 as the main intermediate indicating NSP7 cleavage, which was found at the lower *m*/*z*. In a third step, the polyprotein and intermediate signals almost disappeared showing depletion of all cleavage sites. Delayed NSP8-9 cleavage was further indicated by the rise of NSP9, which at last became the most intense peak. Therefore, the cleavage activity at the junctions was determined from NSP9-10 over NSP7-8 to NSP8-9 with decreasing activity. The results agree with data obtained from native MS where the same order of cleavage was determined.

However, differences between MALDI and native MS require further discussion. *In vitro* processing for native MS was performed in AmAc but for MALDI we were able to use the more physiological phosphate buffered saline. Therefore, similar results obtained endorse the usage of AmAc since it had no alteration of protein structure changing on M_{pro} specificity. Moreover, the MALDI mass spectrum of complete cleavage showed vast difference in ionization efficiency between the NSPs. In fact, the differences appeared broader than determined for native MS (Section 4.1.5). Altogether, MALDI MS is a suitable method for observing the products of cleavage at several cleavage sites. In terms of suitability of the technique, the lack of ionization efficiency is balanced by the fast measuring time and therefore theoretically high sampling frequency.

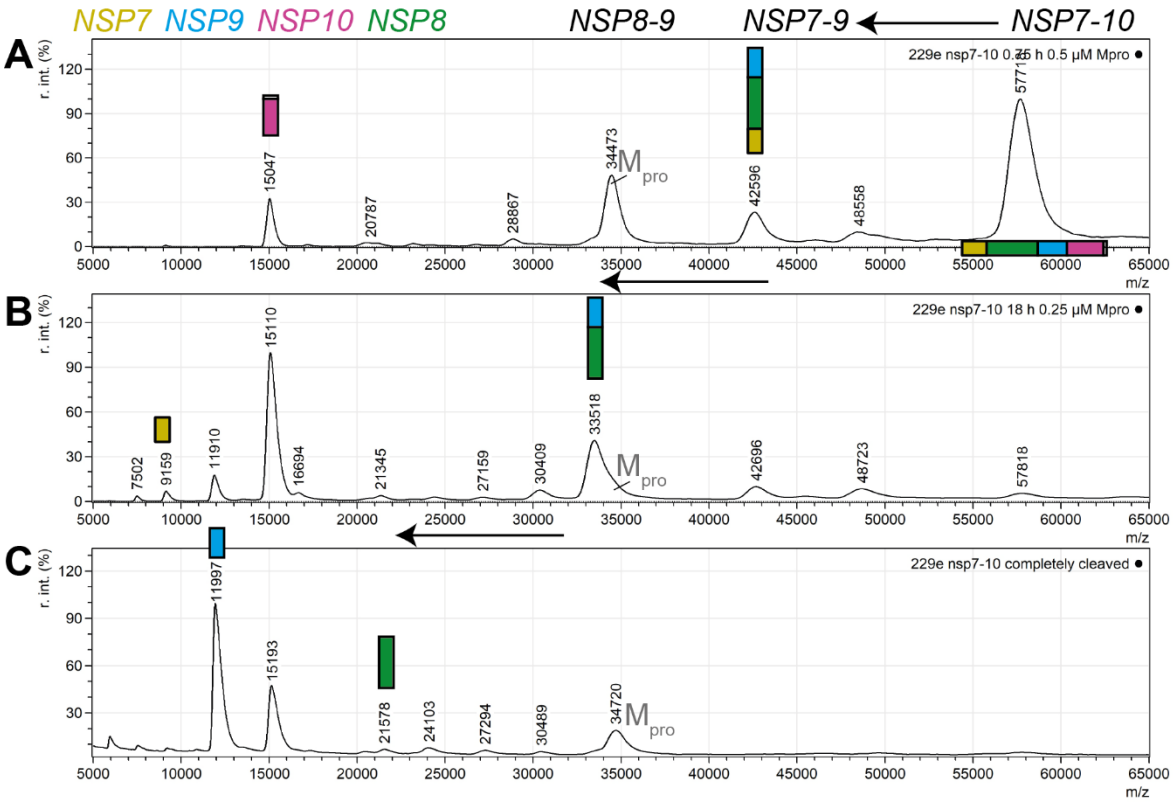


Figure 55: MALDI-MS: *In vitro* processing of 229e NSP7-10. MALDI mass spectra of *in vitro* processing sampled from NSP7-10-His with SARS M_{pro} in 20 mM phosphate buffer, 150 mM NaCl, 1 mM DTT at pH8. Before measurement the samples were diluted to 1 µM, immediately mixed with an acidified sinapinic acid matrix solution and then dried on a MALDI spot plate. Mass spectra show cleavage products in different intensity at three different time points. Arrows indicate the concluded order of processing. **(A)** 10 µM 229e NSP7-10-His + 0.5 µM M_{pro} (1:20) at 45 min shows NSP7-10 polyprotein as well as NSP10 and NSP7-9 as most intense cleavage products. **(B)** 10 µM 229e NSP7-10-His + 0.25 µM M_{pro} (1:40) at 16 h shows polyprotein depleted and NSP8-9 as most intense cleavage intermediate demonstrating that the inter-domain junction is not efficiently cleaved. **(C)** 10 µM 229e NSP7-10-His + 0.5 µM M_{pro} (1:20) at 20 h showing all intermediates depleted. NSP9 relative intensity increased strongly due to cleavage of NSP8-9. NSP monomers have vastly different ionization efficiency. Colored symbols depict polyprotein domains NSP7 (yellow), NSP8 (green), NSP9 (blue) and NSP10 (pink) with highest signal increase. Mass spectra were not calibrated. Each spectrum averaged from triplicate measurements. SARS was chosen because it was readily available and here, the method itself was tested for these experiments. For the 229e substrate the SARS protease is non-natural, which could have implications on cleavage order. However, 229e and SARS M_{pro} have virtually conserved substrate specificities and a significant difference is not expected.

To study complex stoichiometry with MALDI MS, the non-covalent interactions of NSP7+8 required stabilization by chemical cross-linkers. Sample proteins of FIP and 229e were recombinantly produced, affinity purified and *in vitro* cleaved by M_{pro}, and then NSP7+8 complexes were extracted by size exclusion chromatography. To stabilize the complexes chemically, they were incubated with a carefully adjusted concentration of glutaraldehyde, which is a rather unspecific cross-linker. MALDI from a dried sample-matrix spot revealed the stoichiometric landscape of cross-linked NSP7+8.

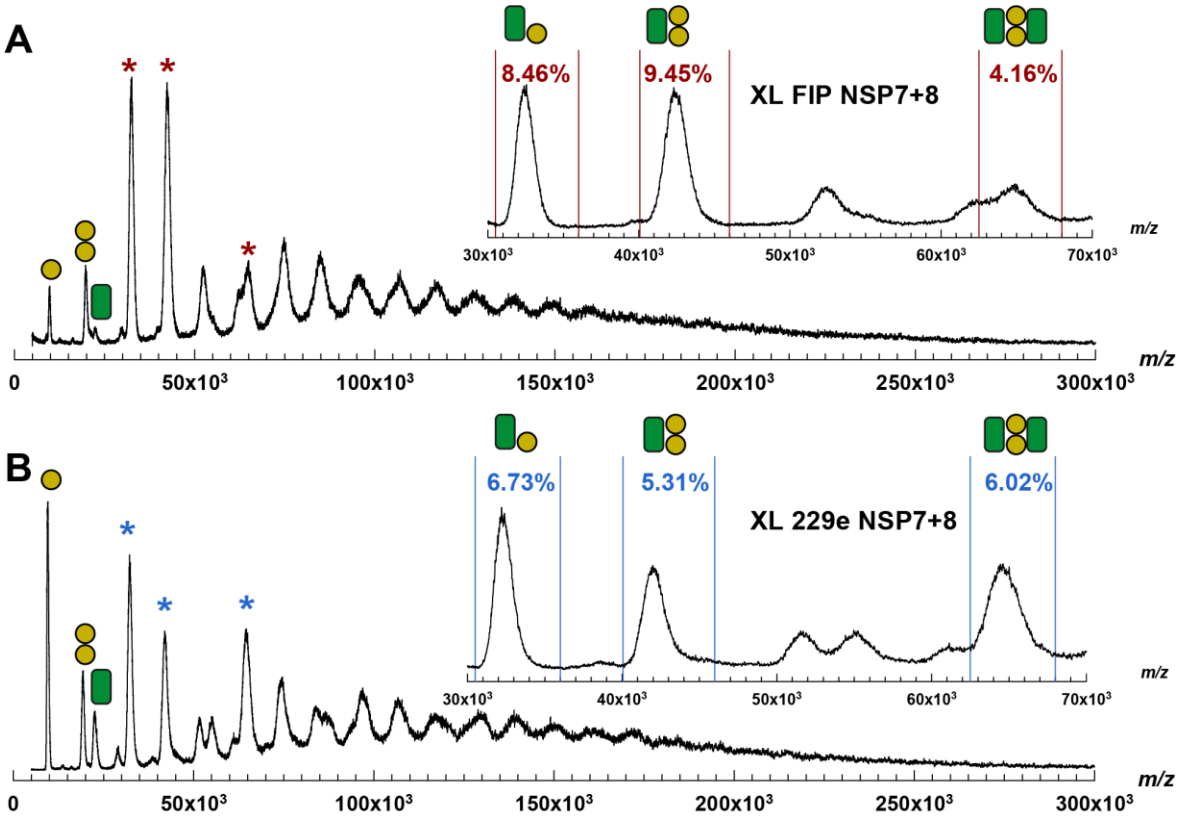


Figure 56: MALDI-MS: NSP7+8 complexes from FIP and 229e stabilized with cross-linker. Complexes purified by size-exclusion chromatography and stored in 20 mM HEPES, 150 mM NaCl, 4 mM ß-mercaptoethanol NSP7+8 in 20 mM HEPES, 150 mM Tris, pH8 were adjusted to 20 µM (~0.6 mg/mL). Before measurement the samples were diluted to 1 µM, immediately mixed with an acidified sinapinic acid matrix solution and then dried on a MALDI spot plate. **(A**) MALDI mass spectrum of FIP NSP7+8 0.15 % Glutaraldehyde for 25 min at 4°C. Inset shows most abundant complexes as (1:1) hetero-dimer (32.4 kDa) and (2:1) hetero-trimer (42.5 kDa), determined from relative peak area as indicated (red). Signals above 50,000 *m/z* most likely due to unspecific crosslinking. **(B)** MALDI mass spectrum of 229e NSP7+8 0.15 % Glutaraldehyde for 25 min at 4°C. Inset shows most abundant complexes stabilized by cross linker as (1:1) hetero-dimer (32.2 kDa) and (2:2) hetero-tetramer (64.6 kDa), determined from relative peak area as indicated (blue). Symbols depict stoichiometric mass species of NSP7 (yellow) and NSP8 (green) with highest signal strength. Deriving masses in cross-linked samples due to weight of the glutaraldehyde molecules. Mass spectra were not calibrated. Each spectrum shown summarized from three MALDI spots.

Without crosslinking, FIP and 229e samples gave most intense signals for NSP7 and NSP8 as singly charged monomers (Figure S 13). In terms of ionization efficiency, both domains were similar, therefore not suggesting an excess of either domain. Then, upon stabilization of the complexes by cross-linking, MALDI spectra showed a variety of higher *m/z* peaks (Figure 56). Analyzing peak intensity appeared unsuitable, due to peak broadening and absence of Gaussian distributions. To consider peak broadening, relative peak areas were estimated by simply setting upper and lower limits. Thereby, the most intense mass species were assigned to (1:1)dimers and (2:1)trimers of FIP NSP7+8, and (1:1)dimers, (2:1)trimers, and (2:2)tetramers of 229e NSP7+8.

In FIP, signals for dimer and trimer were approximately even while in native MS a dimer was present but the (2:1) trimer was the preferred species. Considering the tight binding of NSP8 to NSP7I as indicated by the available structures, NSP8 was more likely cross-linked to NSP7I than to NSP7II. Therefore, the difference in dimer to trimer ratio between native MS and MALDI occurred, at least partly, due to incomplete crosslinking, which then resulted in over-proportioned signal of dimer.

In 229e, similar as in native MS, the tetramer was one of the most abundant mass species. However, the monomeric NSP7 and NSP7:8(1:1) suggested incomplete crosslinking of some domains and therefore an overrepresentation of monomer, dimer and trimer complexes.

From both samples, NSP7(2) dimer was found, which could result either from the dimeric species in solution or from its interaction within the complex. Nevertheless, considering the usage of an nonspecific cross linker in both samples the stoichiometric landscape from MALDI closely resembled observations made in native MS.

To evaluate the usage of MALDI MS for NSP complex formation, some observations should be further discussed. For example, higher *m*/*z* signals in MALDI spectra indicated the presence of many NSP7+8 stoichiometries, which is unlikely. These signals must result from unspecific intermolecular cross-links, which are especially common upon usage of highly reactive chemicals such as glutaraldehyde. Careful optimization of time and cross-linker concentration could reduce these artifacts. Furthermore, cross-linkers with determined length and specific reactivity could help to study side chain placement of complexes with unknown structures, such as 229e tetramer. Finally, pre-purified NSP7+8 complexes could not be buffer exchanged into the native MS AmAc solution, but crosslinking enabled studying their complex formation in a physiological buffer.

Altogether, the dynamics of MALDI-MS could not compete with native MS but proved to be a valuable addition for screening and studying processing and complex formation. To dig even deeper into the dynamical changes from polyprotein to complex, another technique was desired that is able to reveal the structural dynamics of NSP complex formation.

4.5 HDX reveals dynamics from polyprotein to NSP7+8 complex

In this fourth and last set of experiments, hydrogen-deuterium exchange-mass spectrometry (HDX-MS) was used as a bottom-up approach for protein dynamics and therefore was fundamentally different from other techniques used during course of this thesis. For our group, Jasmin Dülfer had established HDX-MS and together, we developed an approach for testing the dynamics of polyprotein NSP7-8 in comparison to complex NSP7+8. In the following paragraphs, I report findings from our initial experiments showing changes in conformation upon complex formation.

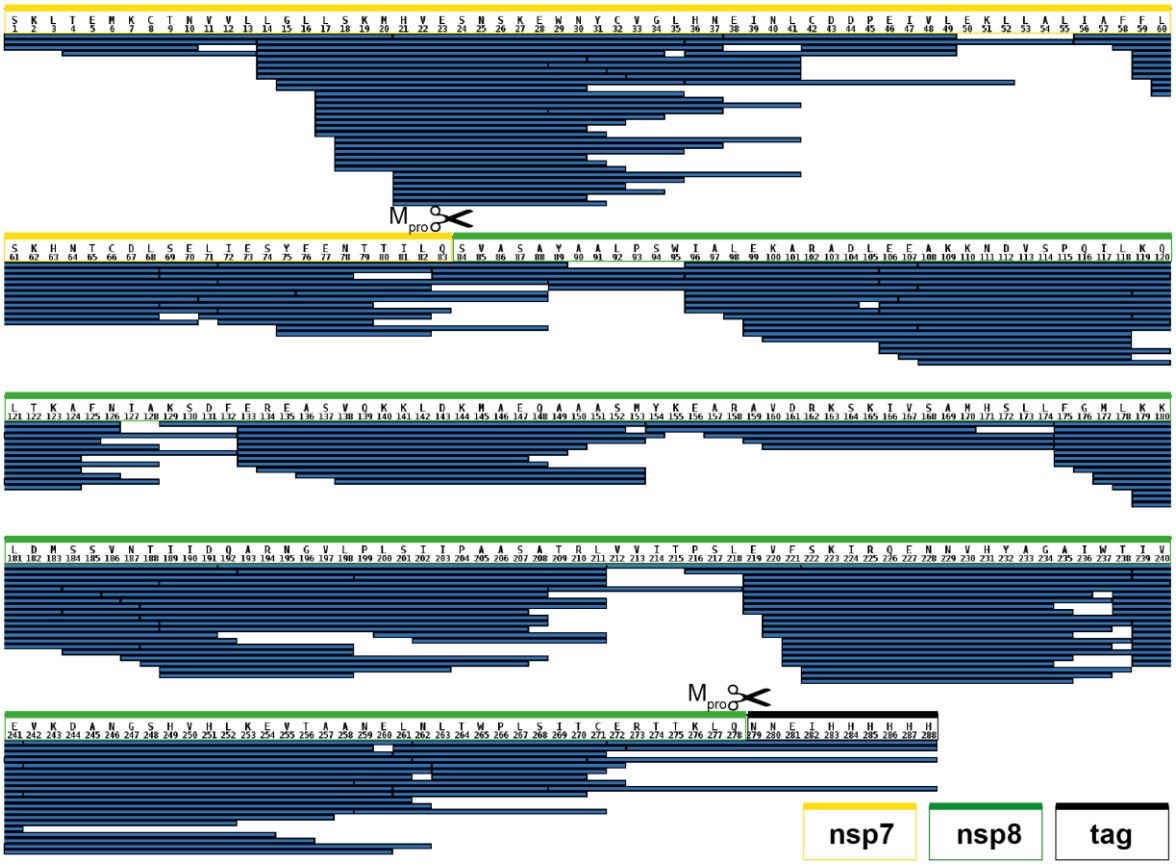


Figure 57: Sequence coverage upon pepsin digestion TGEV NSP7-8. Found peptides of pepsin digested NSP7-8-NNE-His. Sequence coverage of identified peptides is 100%. For HDX data analysis the peptides overlapping with the cleavage sites were excluded. Cleavage sites and products indicated by scissor symbols and color code: NSP7 (yellow), NSP8 (green), and NNE-His tag (black).

For testing the HDX approach, sample protein TGEV NSP7-8-NNE-tag was used. This protein was particularly suited for this technique because upon cleavage a stable trimeric NSP7+8 complex formed. Furthermore, a homology model was generated and confirmed (Figure 45) that could be used for mapping of deuterium uptake. To test peptide coverage, the uncleaved protein was digested in an on-line approach coupled to an ESI-MS readout (Figure 57). Sequence coverage was 100 % of the amino acid sequence, which, in principle, allowed for analyzing deuterium uptake over the complete protein.

To represent polyprotein and complex, TGEV NSP7-8 was used uncleaved and cleaved by M_{pro}, assuming the latter formed complexes as observed in native MS (Figure S 14). The samples were deuterated for 1 min, 10 min or 1 h, quenched and then flash frozen until MS measurement. For analysis, peptides from cleaved and uncleaved sample were compared differentially regarding deuterium uptake (Eq. (28)). Peptides spanning the cleavage site were ignored due to their absence in fully cleaved proteins.

Due to performing only a single measurement with the newly developed approach, criteria for relevant uptake differences had to be set. In the HDX community, as a common rule of thumb, an uptake difference of ±0.5 Da is considered relevant [207]. Therefore, it appeared reasonable, that here uptake was considered relevant when the difference was at least ±0.5 Da and overlapping peptides showed the same trend. Applying these criteria, 45 peptides in eight amino acid regions were identified that show relevant uptake difference of at least ±0.5, and were distributed over NSP7 and NSP8 within cleaved and uncleaved samples (Table 9).

Table 9: Peptides identified with a relevant deuterium difference. Identifier (ID) given to regions that show relevant uptake difference and used for visualization on the molecular model. Domain given for location of the peptide on NSP7 or NSP8. Amino acid region shows peptide coverage using position numbering as in the NSP7-8-NNE-precursor. The difference was determined as follows: uptake difference = uptake[cleaved] – uptake[uncleaved]. Differential uptake is either positive or negative, here tied to lower uptake in either polyprotein or complex, respectively (Exemplary uptake plots: Figure 58 C and D). Number of peptides indicate the number of peptides showing similar relevant differences identified in the depicted region.

region ID	Domain	Amino acid region	Differential uptake ±1	# Peptides
I	NSP7	1-13	-	1
II	NSP8	175-211	-	15
III	NSP8	219-241	-	6
IV	NSP8	241-258	-	1
V	NSP7	14-41	+	15
VI	NSP7	76-88	+	1
VII	NSP8	96–132	+	5
VIII	NSP8	239-259	+	1

The deuterium uptake difference of a peptide between two conditions means that uptake is higher in one condition or lower in the other one. This could result from various dynamic processes in both conditions. However, lower uptake can be interpreted, as that the peptide became more protected, such as in a binding event. To discuss the dynamics of complex formation, the results were visualized on an *in silico* homology model of TGEV (2:1) trimer (Figure 57 A-D).

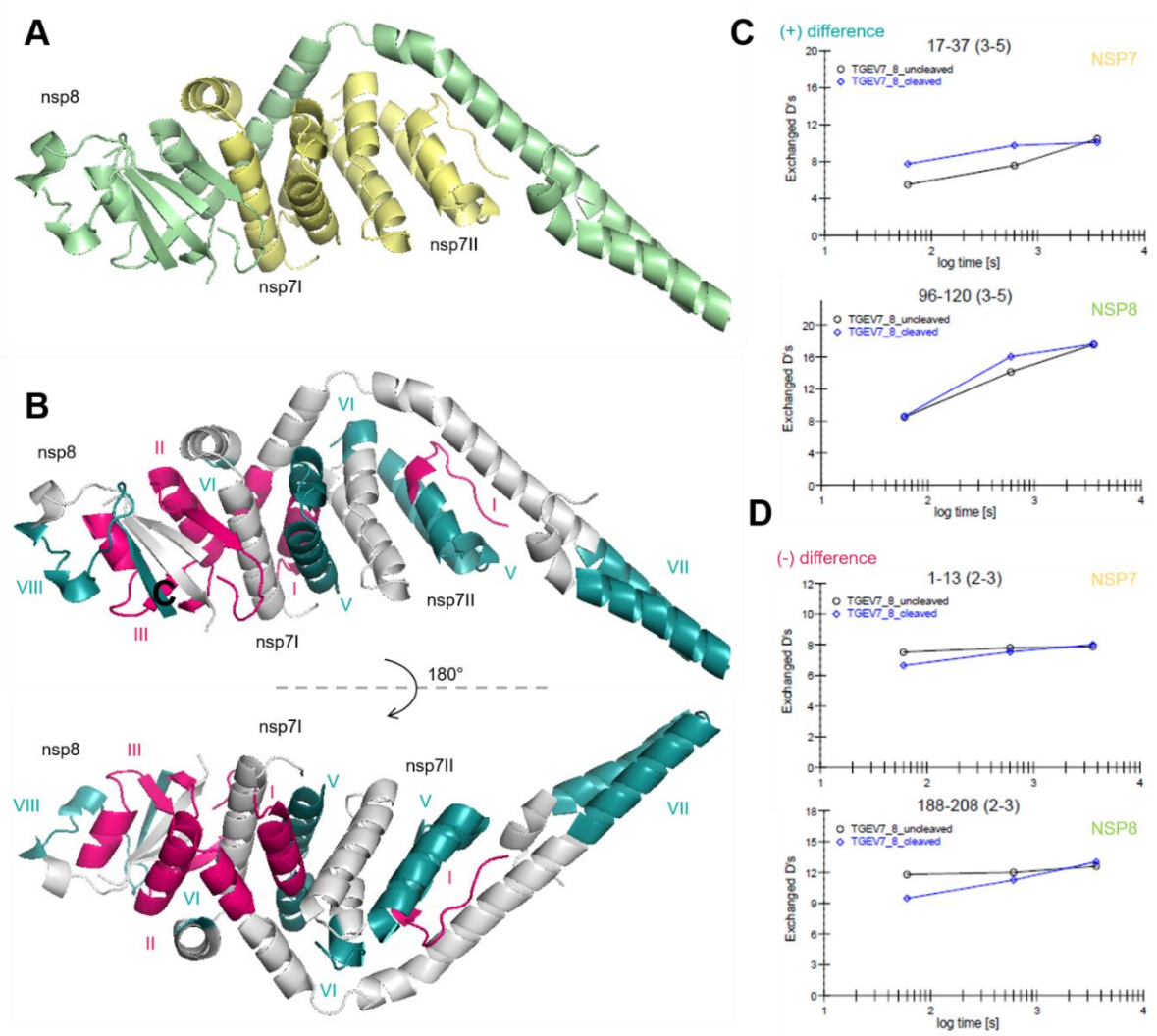


Figure 58: HDX: Structural differences from TGEV NSP7-8 polyprotein to NSP7+8 complex. (A) *In silico* homology model of TGEV (2:1) trimer (Figure 45, PyMOL). Domains indicated as NSP7 (paleyellow).and NSP8 (palegreen) **(B)** Uptake differences visualized on the TGEV model. Deuterium uptake highlighted for positive difference (deepteal), which means a lower uptake in polyprotein, and negative difference (pink), which means a lower uptake in the complex. Region ID is indicated (Table 9). Model rotated 180°. **(C and D)** Selected examples of uptake plots meeting the set criteria for relevant uptake difference between cleaved and uncleaved.

Several differences were interpreted as local structural event in the polyprotein. Most importantly, the regions NSP7 (ID V) and N-terminal NSP8 (ID VII) show a considerably lower uptake in the uncleaved sample. These regions are in proximity of each other in the polyprotein and therefore, the findings indicated for that domains of NSP7 and NSP8 interact in the

polyprotein. Such a higher order interaction would be beneficial, since the elongated NSP8 N-terminus is rather unstable and NSP7 binding might support its fold in a chaperone-like manner (Figure 59 A).

Further, a lower uptake was found in the NSP8 (ID VIII) region of the C-terminal head domain, close to where NSP7I and NSP8 interaction takes place in the complex. Therefore, uptake difference could be due to an allosteric effect in complex formation. However, in the polyprotein, it is intriguing to think about an interaction of NSP7-8 C-terminus with its N-terminus as a folded-up protein, hiding its elongated and aggregation prone form and thus convey stability until it opens up immediately after cleavage. Such a hypothetical fold could also influence substrate efficiency of NSP7-8 cleavage.

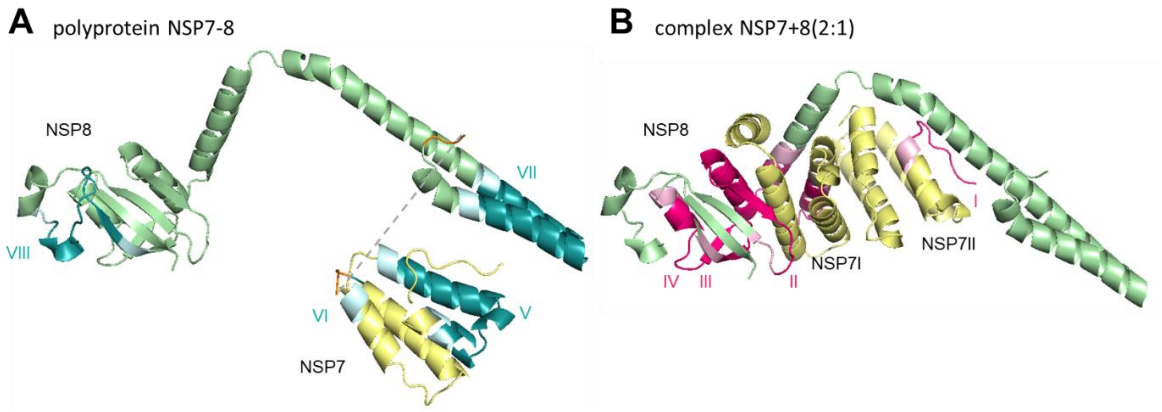


Figure 59: HDX-results illustrated separately for lower uptake in polyprotein and complex. Used data similar as shown in Figure 58. **(A)** Domains of TGEV NSP7 and NSP8 stylistically rearranged as they are most likely found in the polyprotein. C-terminus of NSP7 and N-terminus of NSP8 highlighted (orange), dashed line represents covalent bond between the domains. Highlighted (teal) are regions of lower uptake, which could be interpreted as local binding event. **(B)** Model structure of NSP7+8(2:1). Highlighted (pink) are regions of lower, which could be interpreted as local binding event. The three first and last amino acids within a relevant amino acid region highlighted in a lighter shade color. Region ID indicated (Table 9).

For the trimeric protein complex, three regions of NSP8 were identified with a considerably lower deuterium uptake (ID II, III and IV) (Figure 59 B). All of the peptides were part of the shaft region in NSP8 that connects the head domain with the elongated domain. The shaft region is also the interaction site for NSP7I. Therefore, the differential uptake in the shaft domain was unequivocally based on binding of NSP7I to NSP8 and the local change in structure triggered by this event.

Merely one peptide of NSP7 had lower uptake in the trimeric complex. The structural model shows this peptide (ID I) in the hydrophobic interface between NSP7II and the NSP8 elongated domain. Nevertheless, these interactions in the trimer were not completely verified, probably due to inherent characteristics of the method itself. HDX cannot distinguish between molecules of identical sequence and further, differences of NSP7I and NSP7II could be too low for significant differences.

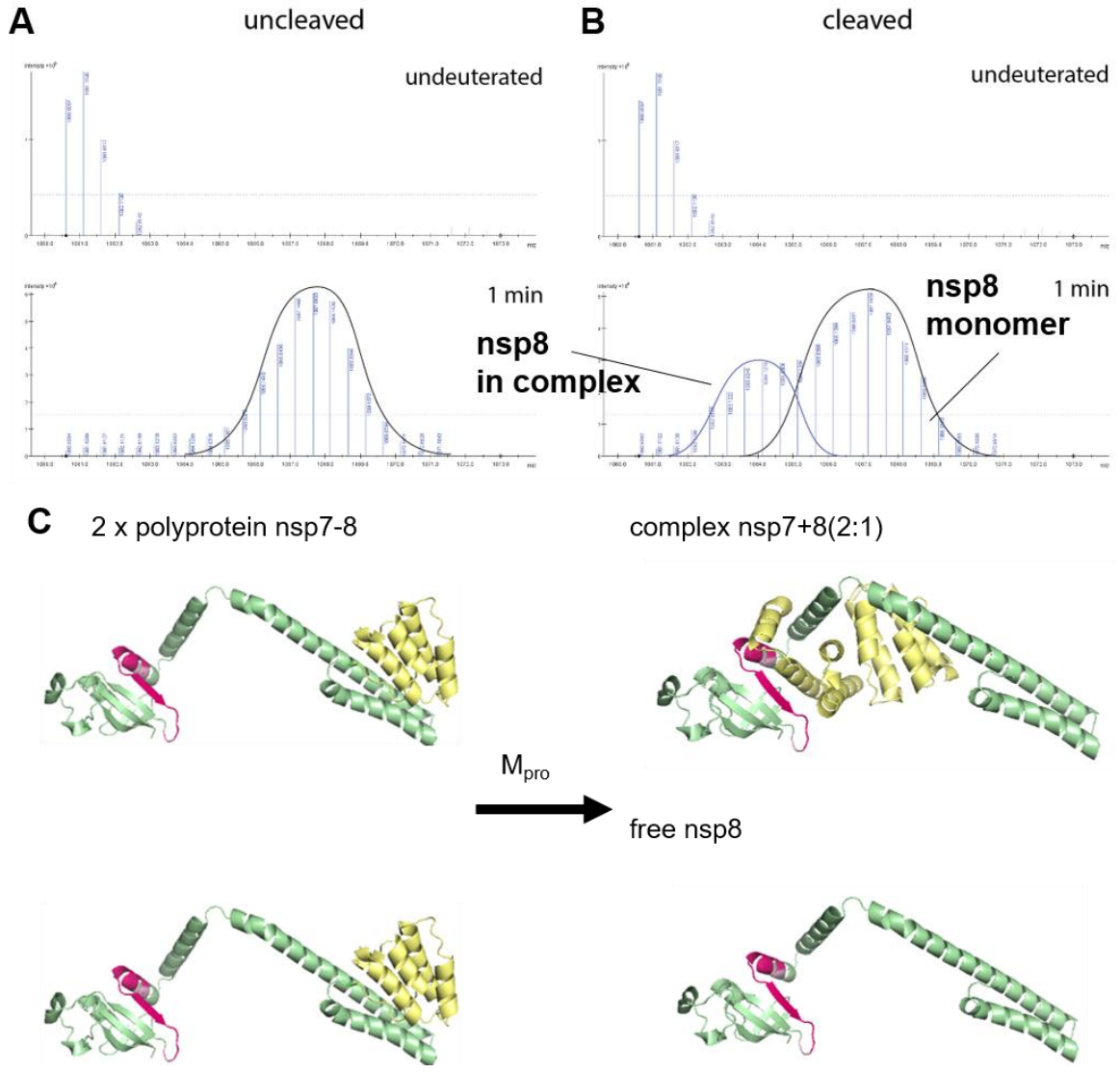


Figure 60: HDX: Two conformers of NSP8 upon complex formation. (A) Peptide aa188-208 of NSP8 from uncleaved NSP7-8-NNE-His un-deuterated (top) and deuterated for 1 min (bottom) shifted unimodal peak distribution. **(B)** Same peptide from cleaved NSP7-8-NNE-His un-deuterated (top) and deuterated for 1 min (bottom) showing bimodal peak distribution. The higher intense peak distribution shows deuterium uptake similar to the uncleaved sample, while the less intense peak distribution shows reduced deuterium uptake. **(C)** Explanatory illustration of polyprotein processing and complex formation, leading to two states of NSP8. Conversion from TGEV NSP7-8 polyprotein to trimeric NSP7+8(2:1) complex generates an excess of monomeric NSP8, which cannot be saturated by NSP7 binding. Therefore, the shifted unimodal peak distribution represent the amino acids aa188-208 (pink) buried in the complex and free in the monomer.

A more detailed analysis of the mass spectra revealed a bimodal deuterium uptake curve in at least one peptide of NSP8 (aa188-208) of the complex after 1 min of deuteration (Figure 60 A-B). In general, bimodal uptake curves indicate structural inhomogeneity for example between two different conformers of one protein.

In fact, this specific NSP8 peptide showed rearrangement from polyprotein to the (2:1) trimeric complex, which by definition ends with an under-representation of the NSP8 domain (Figure 60 C). Since upon cleavage of NSP7-8 both domains must have been present at equimolar concentration, formation of the trimer resulted in at least half the NSP8 domains to be monomeric. Therefore, the bivalent uptake represents complexed and monomeric NSP8, of which the latter most likely had higher deuterium uptake.

Altogether, HDX-MS is indeed a suitable method to generate new information that nicely complements the complex stoichiometry as determined earlier. The presented results gave an even more complete understanding of the dynamics upon structural rearrangement from polyprotein to protein complex. Nevertheless, further measurements are required to obtain statistical security and to confirm these insightful results.

In the following chapter, the conclusions drawn from structural mass spectrometry of non-structural proteins are further discussed and put in context of the virological relevance.

5 Conclusion and outlook

5.1 Sample preparation of NSPs for native MS

The production of NSP proteins appeared to be straightforward in terms of methodology and success. Relatively high amounts of protein were purified and sample analysis revealed proper folding as indicated by many parameters such as stability, mass-to-charge, accurate mass and confirmation of zinc binding.

However, some challenges had to be tackled in order to provide stability to NSP8. In native MS of the 229e NSP7-10 precursor substrate, no unusual charge states, resolution or intensity were observed, but upon processing, NSP8 monomers, complexes or intermediate products having NSP8 at its C-terminal domain were found underrepresented. If NSP8 lost its stability within the polyprotein or immediately after cleavage at its NSP7-8 site, remained unclear. The possibility exists that the relative cleavage efficiency in 229e NSP7-10 was influenced by this instability of the NSP8 domain.

Furthermore, in native MS of complex formation, a disproportionate loss of NSP8 was frequently observed. In one of the later performed experiments in this thesis, the AmAc in the ESI solution was increased from 300 mM to 500 mM and complexes for 229e and PEDV NSP7+8 were identified as predominant mass species, which was a strong indication for increased stability of NSP8. In general, 500 mM AmAc equals a higher ionic strength than what is expected in the cell. Possibly, NSP8 requires chaperones or post-translational modifications to remain stable. Unfortunately, increased AmAc and its influence on processing and complex formation were not tested for all the NSPs and should be addressed in follow-up experiments.

In one prepared sample, the elongated domain of PEDV NSP8 was found to be a target of specific cleavage during sample preparation by an unknown protease(s). This remained a one-time event and its reasons were not further investigated. Other researchers observed such an event before [87] and tried to trigger its generation with a caspase, a protease facilitating cleavage after an aspartic acid. Here, the range of responsible proteases was specified since the cleavage site was located C-terminal of an Arginine residue. However, many cellular protease could facilitate such a reaction. In the future, limited proteolysis could be carried out in order to find the respective product and shed light on this possibly relevant mechanism or chromatographic fractions could be assayed to directly identify this protease.

5.2 Structural MS to analyze polyprotein processing

Protease assays with cleavage site analogue peptides were carried out here and previously by others [60, 82] to evaluate the specific efficiency of processing of polyproteins. While such an assay is acceptable to find an ideal peptide for a protease, it appears critical to compare cleavage efficiency of a peptide to that of a full-length folded protein. In this thesis, I showed that a

cleavage site that was a poor substrate in a FRET peptide was in fact cleaved relatively efficiently in the context of a polyprotein.

Different methods were exploited as readouts for full-length protein processing. In ESI-MS, the challenge remains to correlate signal response to concentration in solution. In the main experiment of this thesis, processing of SARS NSP7-10 with native MS, I solved this task by calculating mass fractions that correct for bias from non-covalent interactions and thereby showed that the assigned peaks in the mass spectra were evenly represent in the cleavage products.

Nevertheless, the used approach is not always valid and could be optimized. One of the challenges was that mass species of NSP complexes gave increased signals compared to NSP monomers. Therefore, the analysis of SARS NSP7-10 was dependent on complex inhibiting fusion tags. These prevented the appearance of larger mass species and ensured final products within a similar *m*/*z* range, which could be compared directly.

The reason for the higher intensity of larger products was that the used mass spectrometer, Q-TOF2, had been modified to effectively transfer high *m*/*z* ions, at the cost of small *m*/*z* ions. It would be interesting to see if the signal response in another instrument would be in better equilibrium in this regard. Furthermore, to calibrate for transmission differences, a calibration mixture of fused maltose binding proteins (MBP) of different length could be used. Such a calibration mix provides ions from 40,000-120,000 Da with similar ion efficiencies, as reported previously by others [212].

In data analysis of the processing experiments, peak intensity was used as a measure for signal response. I showed that ions from similar *m*/*z* range of high quality spectra can be compared and that the results represent the equimolar concentration ion solution. However, inherent isotope distribution or adducts increase the chance of peak broadening and introduce bias. Using peak area instead of intensity would correct for these factors. Nevertheless, the analysis of Gaussian peak fitting by hand would have been out of proportion, since a single analysis of processing with six time-points in triplicate measurements resulted in more than 500 peaks. Software exists about that matter and I have used Massign [203] to fit peaks in envelopes and calculate peak areas for a smaller and more static system [213]. However, available software is far from being useable for large datasets with dynamically increasing and decreasing peak envelopes.

In order to determine absolute efficiencies of cleavage sites in full-length proteins, particularly NSP-NSP, substrate with only a single inter-domain junction is required. However, in the here conducted ESI-MS experiments, the analysis was hindered by cleavage sites of artificial tags

and by the relatively low sampling rate. In nanoESI, a maximum of about one high quality spectra per 5-10 minutes is possible, limited by mounting a new sample on the source. An automatic sampling device could ensure a more stable quality of ESI capillaries and a higher sampling rate. With MALDI-MS, a higher sampling rate could be possible since the MALDI sample spots can be prepared independent of the analysis. Furthermore, the MALDI mass spectra are simpler to analyze, since only one singly charged ion is predominant in the mass spectrum. Notably about the general setup of the assay is that a mixture of NSP-NSP substrates in a cleavage assay could directly reveal the relative efficiency of a substrate.

5.3 Influence of polyprotein structure on processing

The structure of polyproteins is an important factor for cleavage site accessibility and thus a layer of regulation for processing. Higher order structures can also influence cleavage efficiencies and consequently, a change in relative efficiencies between the virus species could indicate that arrangement is different. Some experiments shed light on the question of if the CoV polyprotein is an array of folded domains connected by flexible cleavage sites like 'pearls on a string' or an overall folded entity with higher order features. The results in this thesis suggested a mixture of these models. On the one hand, the repeated observation of efficient cleavage of the NSP9-10 site, both in peptides as well as in full-length substrates, is rather indicative of a stretched out or flexible linker between NSP9 and NSP10 in the context of the polyprotein. On the other hand, the HDX-MS results suggest extensive interactions between NSP7 and the N-terminal domain of NSP8 in the polyprotein, and therefore support that higher order interactions exist.

Remarkably, the interaction of NSP8 and NSP7 in the polyprotein would represent a third NSP7 binding site on the C-terminal domain of NSP8, next to the NSPI and NSPII binding sites. A third interaction site could be necessary to convey stability to the NSP8 in the polyprotein context. This hypothesis agrees with the identification of the NSP8 N-terminal domain as a factor for instability and aggregation immediately upon cleavage from NSP7. Additionally, it could explain why other researchers reported that switching domains from NSP7-NSP8 to NSP8-NSP7 led to unviable mutants. To collect more evidence, a recombinant protein with switched domains could be analyzed with native MS and HDX-MS to assess stability, binding sites and cleavage efficiency.

Processing of different coronavirus substrates was successful, regarding developing the approach and observing order of processing of different substrates. Analyzing full-length substrates with Native MS gave more conclusive results than peptide substrates did. However, even though the NSP9-10 site had a relatively high cleavage efficiency in SARS and 229e substrates, the sample size was too small to draw general conclusions about conservation of cleavage efficiency.

The relative cleavage efficiency of the NSP8-9 site was particularly interesting since its amino acid motif is non-canonical but generally conserved among coronaviruses with a strictly conserved NNE motif (P'1-P'3). Results from full-length protein substrates show differences between NSP8-9 of SARS and 229e regarding relative cleavage efficiency. However, between these viral species' NSP8-9 sites, the amino acids are conserved or conservatively replaced at P5-P4'. Therefore, it appears unlikely that the conserved NNE motif leads per se to inefficiency of the cleavage site. Moreover, a functional intermediate protein NSP8-9 cannot exist for a prolonged time as first, it would not be favored by the efficiencies detected here and second, it would most likely be unstable due to the elongated domain of NSP8 requiring a binding partner. Hence, the results suggest that the strictly conserved NNE motif in the N-terminus of NSP9 has another biological function that remains elusive. Since the sites for dimerization and amino acid binding in NSP9 were already identified, biochemical assays seem to be not the right choice to screen for a possible function of this motif, which has to be tackled in cellular assays.

Of course to fully understand processing of M_{pro}, an ideal assay would use a membrane anchored full-length NSP4-16 polyprotein. Even though membrane proteins can be analyzed with native MS, such an assay would require extensive sample preparation that appears challenging or even unrealistic. However, some samples could be produced relatively easy. For instance, an extended NSP7-10 domain with a soluble N-terminal NSP6c and C-terminal NSP11 could be used. Such NSP6c-NSP11 protein substrate could act as a binding scaffold to test protein-protein interactions with core enzymes and their influence on processing. It would be particularly interesting to know if NSP14 and NSP16 are able to build NSP10+14/NSP10+16 complexes before NSP10-11 is cleaved and if these complexes could fulfill their enzymatic function with similar efficiency.

5.4 Role of the observed complexes in the replication-transcription complex

In this thesis, I acquired a comprehensive picture of all interactions between the domains of the NSP7-10 region, which are a requirement for the coronaviral RTC. Amongst these interactions was the intermediate complex 229e NSP7-9+NSP7(1:1). Both of its constituents were temporally enriched by coordinated cleavage of 229e NSP7-10. Hence this complex was a direct result of processing. Current knowledge suggests that NSP7 in this complex interacts with the NSP8 domain of NSP7-9. Therefore, this structure represents a class of coronavirus complexes that has not been yet observed, namely the binding of a matured NSP to an unprocessed polyprotein precursor. Polyprotein to NSP complexes have been discussed as a switch to tweak different functions in replication. To find the maximal complex 'load', it could be revealing to reenact a possible scenario during viral replication by adding an excess of matured NSPs to unprocessed precursor NSP7-10.

The analysis of NSP7+8 from SARS, TGEV, FIP, PEDV and 229e showed various stoichiometries and subunit arrangements. However, a sequence alignment did not allow comprehensive conclusion about the features determining trimer or tetramer formation. Up to now, this sequence analysis appears to not be powerful enough to use the results, merely two decisive residues, for mutational analysis. If NSP7+8 from other species would be analyzed in native MS, the information could be added to the existing dataset to provide more conclusive results on which amino acids are involved.

Puzzling is that for NSP7(2) dimer only protomer presents the conserved primary interaction site while the other one is buried in the dimer interface. Therefore, in PEDV and 229e tetramers, a central NSP7(2) dimer should only be able to interact with one NSP8 at its primary binding site. It could be possible that the second NSP8 uses the second interaction site. Such an arrangement would resemble the structure of FIP NSP7+8 trimer. This would also explain why the gas-phase dissociation products of PEDV and 229e tetramer resembled TGEV and FIP trimer.

HDX-MS revealed the peptide dynamics upon processing of TGEV NSP7-8 and rearrangement to NSP7+8 trimeric complexes (2:2). Of course, the here conducted experiments did not provide statistical security of a triplicate measurement. However, preliminary validation was given by distinct differences in deuterium uptake pointing to conserved binding sites for NSP7+8. Therefore, the results allowed a careful interpretation of the acquired data. After all, HDX-MS qualified as a promising technique to clarify open questions. For instance, the differences between the tetramer arrangements of SARS and PEDV/229e. An even clearer picture would be generated by not only comparing uptake differences between uncleaved and cleaved precursor, but also free monomeric proteins.

A glimpse of the fate of NSP7 and NSP8 was given by the cryoEM structure of the SARS polymerase complex NSP12+8+(7+8)(1:1:1) [93]. Some questions were raised by two different binding sites for the monomeric NSP8 and NSP7+8(1:1) hetero-dimer. For example, do the two sites allow a different stoichiometry by binding to the NSP8 monomer and NSP7+8(2:1) trimer in virus species such as TGEV and FIP? Does this use all the generated subunits of NSP7 and NSP8? Furthermore, does such a complex directly assemble at the polyprotein upon cleavage so that NSP12 'picks-up' the NSP7+8 subunits? Since, the results of this thesis suggest that NSP7 is important for the stability of NSP8, trimeric and tetrameric complexes could function as intermediates that impart stability. Therefore, these complexes could also be a necessary requirement for formation of the polymerase complex. Many of these questions could possibly be answered by extending the native MS analysis to NSP12.

To sum up, in this thesis I shed light on different aspects of processing and complex formation of the NSP7-10 region of coronaviruses and thereby showed that these processes are tied to each other. The used methods in structural MS proved well suited for future analysis of a more complete *in vitro* assembly of the coronaviral RTC.

6 Indices

6.1 Index tables

6.2 Index figures

7 Supplement

7.1 Indices supplement

Index supplementary tables

Index supplementary figures

7.2 Molecular mass tables

Table S 1: Molecular weights (MWs) of NSPs, their cleavage products and complexes. Average theoretical mass ($MW_{Theo.}$) according to amino acid sequence (Table S 2). Molecular weight (MW) ($MW_{Exp.}$) determined from Native MS or CID (*) and listed with standard deviations (St.dev.) (N=3) and the average full-width half maximum (FWHM) (N=3) of the main peak. All values in Da. The masses are given without St. Dev. for mass species that were safely assigned in only two spectra.

Protein	Species	MW_{Theo}	$MW_{Exp.}$	St.dev.	FWHM
SARS His-NSP7-10 processing					
His-NSP7	SARS	10677	10680.1	0.2	1.4
NSP8	SARS	21866	21871	1	1.4
NSP9	SARS	12401	12403.5	0.2	1.3
NSP10	SARS	14843	14974	1	1.9
His-NSP7-8	SARS	32525	32541	11	1.4
His-NSP7-9	SARS	44908	44937	15	3.9
His-NSP7-10	SARS	59734	59930	40	14.4
229e NSP7-10-His processing					
NSP7	229e	9301	9300	1	1.3
NSP8	229e	-	-	-	-
NSP9	229e	12046	12045.2	0.3	1.3
NSP10	229e	15218	15346	5	1.5
NSP7-8	229e	30906	30931	8	1.4
NSP8-9	229e	33651	33657	2	1.5
NSP7-9	229e	42933	42961	19	5.0
NSP7-9+NSP7(1:1)	229e	52216	52260	20	6.1
NSP7-10-His	229e	58135	58286	27	10.0
FIP 7-9-His processing					
NSP7	FIP	-	-	-	-
NSP8	FIP	-	-	-	-
NSP9	FIP	12106	12105.4	0.3	1.4
NSP7-8	FIP	30833	30913.6	2.6	6.9
NSP7+8(2:1)	FIP	40234	40360	10	13.7
NSP7-9	FIP	42921	43034	16	16.8
NSP7-9-His	FIP	43744	43960	40	18.7

Table S 1 continues on the next page.

Table S 1 (continuing): Molecular weights (MWs) of NSPs, their cleavage products and complexes.

Complex	Species	$MW_{Theo.}$	$MW_{Exp.}$	St.dev.	FWHM
SARS complexes					
NSP7	SARS	9267.8	9268.3	0.1	2.4
NSP8	SARS	21866.0	21867.3	0.5	3.7
NSP8(2) dim	SARS	43713.9	43700*	30	16
NSP7+8(2:1)	SARS	40365.6	40404*	8	3.6
NSP7+8(1:2)	SARS	52963.7	53009*	6	4.5
NSP7+8(2:2)	SARS	62213.5	62300	10	7.1
FIP complexes					
NSP7	FIP	9418.8	9425	4	2.8
NSP8a	FIP	21432.8	21448*	11	2.4
NSP8b	FIP	21432.8	21531*	14	3.3
NSP7(2) dim	FIP	18819.6	18834*	2	4.8
NSP7+8(1:1)a	FIP	30833.6	30853.3*	2.8	4
NSP7+8(1:1)b	FIP	30833.6	30937*	12	4.9
NSP7+8(2:1)a	FIP	40234.4	40290	4	2.8
NSP7+8(2:1)b	FIP	40234.4	40369	7	3.7
TGEV complexes					
NSP7	TGEV	9489.0	9497	2	2.0
NSP8	TGEV	21458.8	21415*	2	2.3
NSP7(2) dim	TGEV	18959.9	18977.1*	0.1	4.4
NSP7+8(1:1)	TGEV	30929.7	30950*	10	3.2
NSP7+8(2:1)	TGEV	40400.7	40446	12	3.8
PEDV complexes					
NSP7	PEDV	9224.5	9224.8	0.2	1.9
NSP8	PEDV	21688.77	21692*	6	2.8
NSP7(2)	PEDV	18431.0	1844.5	-	2.2
			18448.9*		2.4
NSP7+8(1:1)	PEDV	30895.3	30928	9	3.7
NSP7+8(2:1)	PEDV	40101.8	40144*	6	6.8
NSP7+8(2:2)	PEDV	61722.6	61850	20	5.8
229e complexes	PEDV				
NSP7	229e	9300.6	9301	-	1.8
			9301		1.5
NSP8	229e	21623.1	21624.2*	0.3	1.6
NSP7(2)	229e	18583.2	18600.3	-	2.8
			18600.1*		2.9
NSP7+8(1:1)	229e	30905.7	30934.0	-	5.3
			30934.4		4.2
NSP7+8(2:1)	229e	40188.2	40220*	10	9.3
NSP7+8(2:2)	229e	61793.3	61861.3	-	3.0
			61869.5		5.0

7.3 Sequences and sequence alignments

Table S 2: Amino acid sequences of pp1a/ab (region NSP4/5-NSP10/11/12). Sequences shown of SARS, TGEV, FIP, PEDV and 229e. NSP domains separated visually. Swiss Prot Protein ID or NCBI Identifier gives source of sequence in the header.

>sp|P0C6X7|R1AB_CVHSA Replicase polyprotein 1ab Human SARS coronavirus
NSP4 C-term
TSAVLQ
NSP5
SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNPNYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRL
KVDTSNPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLE
GKFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCAA
LKELLQNGMNGRTILGSTILEDEFTPFDVVRQCSGVTFQ
NSP6
GKFKKIVKGTHHWMLLTFLTSLLILVQSTQWSLFFFVYENAFLPFTLGIMAIAACAMLLVKHKHAFLCLFLLPSLATVAYFNMVYMPAS
WVMRIMTWLELADTSLSGYRLKDCVMYASALVLLILMTARTVYDDAARRVWTLMNVITLVYKVYYGNALDQAISMWALVISVTSNYSGV
VTTIMFLARAIVFVCVEYYPLLFITGNTLQCIMLVYCFLGYCCCCYFGLFCLLNRYFRLTLGVYDYLVSTQEFRYMNSQGLLPPKSSID
AFKLNIKLLGIGGKPCIKVATVQ
NSP7
SKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINRLCEEMLDNRATLQ
NSP8
AIASEFSSLPSYAAYATAQEAYEQAVANGDSEVVLKKLKKSLNVAKSEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQT
MLFTMLRKLDNDALNNIINNARDGCVPLNIIPLTTAAKLMVVVPDYGTYKNTCDGNTFTYASALWEIQQVVDADSKIVQLSEINMDNSP
NLAWPLIVTALRANSAVKLQ
NSP9
NNELSPVALRQMSCAAGTTQTACTDDNALAYYNNSKGGRFVLALLSDHQDLKWARFPKSDGTGTIYTELEPPCRFVTDTPKGPKVKYLY
FIKGLNNLNRGMVLGSLAATVRLQ
NSP10
AGNATEVPANSTVLSFCAFAVDPAKAYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDHPNPKGF
CDLKGKYVQIPTTCANDPVGFTLRNTVCTVCGMWKGYGCSCDQLREPLMQ
NSP11/12 Nterm
SADAST

>sp|P0C6Y5|R1AB_CVPPU Replicase polyprotein 1ab Porcine transmissible gastroenteritis coronavirus (strain Purdue
NSP4 C-term
VNSTLQ
NSP5
SGLRKMAQPSGLVEPCIVRVSYGNNVLNGLWLGDEVICPRHVIASDTTRVINYENEMSSVRLHNFSVSKNNVFLGVVSARYKGVNLVLK
VNQVNPNTPEHKFKSIKAGESFNILACYEGCPGSVYGVNMRSQGTIKGSFIAGTCGSVGYVLENGILYFVYMHHLELGNGSHVGSNFEG
EMYGGYEDQPSMQLEGTNVMSSDNVVAFLYAALINGERWFVTNTSMSLESYNTWAKTNSFTELSSTDAFSMLAAKTGQSVEKLLDSIVR
LNKGFGGRTILSYGSLCDEFTPTEVIRQMYGVNLQ
NSP6
AGKVKSFFYPIMTAMTILFAFWLEFFMYTPFTWINPTFVSIVLAVTTLISTVFVSGIKHKMLFFMSFVLPSVILVTAHNLFWDFSYYES
LQSIVENTNTMFLPVDMQGVMLTVFCFIVFVTYSVRFFTCKQSWFSLAVTTILVIFNMVKIFGTSDEPWTENQIAFCFVNMLTMIVSLT
TKDWMVVIASYRIAYYIVVCVMPSAFVSDFGFMKCISIVYMACGYLFCCYYGILYWVNRFTCMTCGVYQFTVSAAELKYMTANNLSAPK
NAYDAMILSAKLIGVGGKRNIKISTVQ
NSP7
SKLTEMKCTNVVLLGLLSKMHVESNSKEWNYCVGLHNEINLCDDPEIVLEKLLALIAFFLSKHNTCDLSELIESYFENTTILQ
NSP8
SVASAYAALPSWIALEKARADLEEAKKNDVSPQILKQLTKAFNIAKSDFEREASVQKKLDKMAEQAAASMYKEARAVDRKSKIVSAMHS
LLFGMLKKLDMSSVNTIIDQARNGVLPLSIIPAASATRLVVITPSLEVFSKIRQENNVHYAGAIWTIVEVKDANGSHVHLKEVTAANEL
NLTWPLSITCERTTKLQ
NSP9
NNEIMPGKLKERAVRASATLDGEAFGSGKALMASESGKSFMYAFIASDNNLKYVKWESNNDIIPIELEAPLRFYVDGANGPEVKYLYFV
KNLNTLRRGAVLGYIGATVRLQ
NSP10
AGKPTEHPSNSSLLTLCAFSPDPAKAYVDAVKRGMQPVNNCVKMLSNGAGNGMAVTNGVEANTQQDSYGGASVCIYCRCHVEHPAIDGL
CRYKGKFVQIPTGTQDPIRFCIENEVCVVCGCWLNNGCMCDRTSMQSFTVDQ
NSP11/12 Nterm
SYLNRV

sp|Q98VG9|R1AB_FIPV Replicase polyprotein 1ab OS=Feline coronavirus
NSP4 C-term
VNSTLQ
NSP5
SGLRKMAQPSGVVEPCIVRVAYGNNVLNGLWLGDEVICPRHVIASDTSRVINYENELSSVRLHNFSIAKNNAFLGVVSAKYKGVNLVLK
VNQVNPNTPEHKFKSVRPGESFNILACYEGCPGSVYGVNMRSQGTIKGSFIAGTCGSVGYVLENGTLYFVYMHHLELGNGSHVGSNLEG
EMYGGYEDQPSMQLEGTNVMSSDNVVAFLYAALINGERWFVTNTSMTLESYNAWAKTNSFTEIVSTDAFNMLAAKTGYSVEKLLECIVR
LNKGFGGRTILSYGSLCDEFTPTEVIRQMYGVNLQ

NSP6
SGKVKSIFYPMMTAIAILFAFWLEFFMYTPFTWINPTFVSVVLAITTLVSVLLVAGIKHKMLFFMSFVMPSVILATAHNVVWDMTYYES
LQVLVENVNTTFLPVDMQGVMLALFCVVVFVICTIRFFTCKQSWFSLFATTIFVMFNIVKLLGMIGEPWTDDHFLLCLVNMLTMLISLT

TKDWFVVFASYKVAYYIVVYVMQPAFVQDFGFVKCVSIIYMACGYLFCCYYGILYWVNRFTCMTCGVYQFTVSPAELKYMTANNLSAPK
TAYDAMILSFKLMGIGGGRNIKISTVQ

NSP7

SKLTEMKCTNVVLLGLLSKMHVESNSKEWNYCVGLHNEINLCDDPDAVLEKLLALIAFFLSKHNTCDLSDLIESYFENTTILQ

NSP8

SVASAYAALPSWIAYEKARADLEEAKKNDVSPQLLKQLTKACNIAKSEFEREASVQKKLDKMAEQAAASMYKEARAVDRKSKIVSAMHS
LLFGMLKKLDMSSVNTIIEQARNGVLPLSIIPAASATRLIVVTPNLEVLSKVRQENNVHYAGAIWSIVEVKDANGAQVHLKEVTAANEL
NITWPLSITCERTTKLQ

NSP9

NNEILPGKLKEKAVKASATIDGDAYGSGKALMASEGGKSFIYAFIASDSNLKYVKWESNNDVIPIELEAPLRFYVDGVNGPEVKYLYFV
KSLNTLRRGAVLGYIGATVRLQ

NSP10

AGKPTEHPSNSGLLTLCAFAPDPAKAYVDAVKRGMQPVTNCVKMLSNGAGNGMAITNGVESNTQQDSYGGASVCIYCRCHVEHPAIDGL
CRFKGKFVQVPTGTQDPIRFCIENEVCVVCGCWLTNGCMCDRTSIQGTTIDQ

NSP11/12 Nterm

SYLNEC

>AFV09313.1 polyprotein 1ab [Porcine epidemic diarrhea virus]

NSP4 C-term

YNSTLQ

NSP5

AGLRKMAQPSGVVEKCIVRVCYGNMALNGLWLGDTVMCPRHVIASSTTSTIDYDYALSVLRLHNFSISSGNVFLGVVGVTMRGALLQIK
VNQNNVHTPKYTYRTVRPGESFNILACYDGAAAGVYGVNMRSNYTIRGSFINGACGSPGYNINNGTVEFCYLHQLELGSGCHVGSDLDG
VMYGGYEDQPTLQVEGASSLFTENVLAFLYAALINGSTWWLSSSRIAVDRFNEWAVHNGMTTVVNTDCFSILAAKTGVDVQRLLASIQS
LHKNFGGKQILGYTSLTDEFTTGEVIRQMYGVNLQ

NSP6

SGYVSRACRNVLLVGSFLTFFWSELVSYTKFFWVNPGYVTPMFACLSLLSSLLMFTLKHKTLFFQVFLIPALIVTSCINLAFDVEVYNY
LAEHFDYHVSLMGFNAQGLVNIFVCFVVTILHGTYTWRFFNTPVSSVTYVVALLTAAYNYFYASDILSCAMTLFASVTGNWFVGAVCYK
AAVYMALRFPTFVAIFGDIKSVMFCYLVLGYFTCCFYGILYWFNRFFKVSVGVYDYTVSAAEFKYMVANGLRAPTGTLDSLLLSAKLIG
IGGERNIKISSVQ

NSP7

SKLTDIKCSNVVLLGCLSSMNVSANSTEWAYCVDLHNKINLCNDPEKAQEMLLALLAFFLSKNSAFGLDDLLESYFNDNSMLQ

NSP8

SVASTYVGLPSYVIYENARQQYEDAVNNGSPPQLVKQLRHAMNVAKSEFDREASTQRKLDRMAEQAAAQMYKEARAVNRKSKVVSAMHS
LLFGMLRRLDMSSVDTILNLAKDGVVPLSVIPAVSATKLNIVTSDIDSYNRIQREGCVHYAGTIWNIIDIKDNDGKVVHVKEVTAQNAE
SLSWPLVLGCERIVKLQ

NSP9

NNEIIPGKLKQRSIKAEGDGIVGEGKALYNNEGGRTFMYAFISDKPDLRVVKWEFDGGCNTIELEPPRKFLVDSPNGAQIKYLYFVRNL
NTLRRGAVLGYIGATVRLQ

NSP10

AGKQTEQAINSSLLTLCAFAVDPAKTYIDAVKRGHKPVGNCVKMLANGSGNGQAVTNGVEASTNQDSYGGASVCLYCRAHVEHPSMDGF
CRLKGKYVQVPLGTVDPIRFVLENDVCKVCGCWLANGCTCDRSIMQ

NSP11/12 Nterm

STDYGL

>NP_073549.1 replicase polyprotein 1ab [Human coronavirus 229E]

NSP4 C-term

YGSTLQ

NSP5

AGLRKMAQPSGFVEKCVVRVCYGNTVLNGLWLGDIVYCPRHVIASNTTSAIDYDHEYSIMRLHNFSIISGTAFLGVVGATMHGVTLKIK
VSQTNMHTPRHSFRTLKSGEGFNILACYDGCAQGVFGVNMRTNWTIRGSFINGACGSPGYNLKNGEVEFVYMHQIELGSGSHVGSSFDG
VMYGGFEDQPNLQVESANQMLTVNVVAFLYAAILNGCTWWLKGEKLFVEHYNEWAQANGFTAMNGEDAFSILAAKTGVCVERLLHAIQV
LNNGFGGKQILGYSSLNDEFSINEVVKQMFGVNLQ

NSP6

SGKTTSMFKSISLFAGFFVMFWAELFVYTTTIWVNPGFLTPFMILLVALSLCLTFVVKHKVLFLQVFLLPSIIVAAIQNCAWDYHVTKV
LAEKFDYNVSVMQMDIQGFVNIFICLFVALLHTWRFAKERCTHWCTYLFSLIAVLYTALYSYDYVSLLVMLLCAISNEWYIGAIIFRIC
RFGVAFLPVEYVSYFDGVKTVLLFYMLLGFVSCMYYGLLYWINRFCKCTLGVYDFCVSPAEFKYMVANGLNAPNGPFDALFLSFKLMGI
GGPRTIKVSTVQ

NSP7

SKLTDLKCTNVVLMGILSNMNIASNSKEWAYCVEMHNKINLCDDPETAQELLLALLAFFLSKHSDFGLGDLVDSYFENDSILQ

NSP8

SVASSFVGMPSFVAYETARQEYENAVANGSSPQIIKQLKKAMNVAKAEFDRESSVQKKINRMAEQAAAAMYKEARAVNRKSKVVSAMHS
LLFGMLRRLDMSSVDTILNMARNGVVPLSVIPATSAARLVVVVPDHDSFVKMMVDGFVHYAGVVWTLQEVKDNDGKNVHLKDVTKENQE
ILVWPLILTCERVVKLQ

NSP9

NNEIMPGKMKVKATKGEGDGGITSEGNALYNNEGGRAFMYAYVTTKPGMKYVKWEHDSGVVTVELEPPCRFVIDTPTGPQIKYLYFVKN
LNNLRRGAVLGYIGATVRLQ

NSP10

AGKQTEFVSNSHLLTHCSFAVDPAAAYLDAVKQGAKPVGNCVKMLTNGSGSGQAITCTIDSNTTQDTYGGASVCIYCRAHVAHPTMDGF
CQYKGKWVQVPIGTNDPIRFCLENTVCKVCGCWLNHGCTCDRTAIQ

NSP11/12 Nterm

SFDNSY

7.4 Supplementary figures

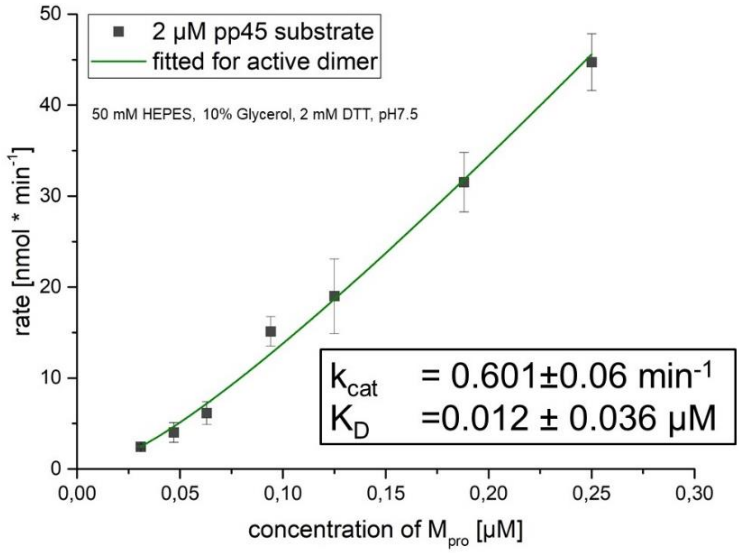


Figure S 1: M_{pro} dimer dependent activity. FRET peptide assay showing activity versus protease concentration. For showing dimer dependent activity a 20 µM FRET peptide substrate FPS4-5 (final concentration 2 µM) was incubated with varying amount of SARS Mpro (1.0 µM, 0.75 µM, 0.5 µM, 0.25 µM, 0.188 µM, 0.125 µM, 0.094 µM, 0.063 µM, 0.047 µM, 0.031 µM and 0 µM). Error bars indicate for standard deviation from triplicate measurements. The green curve is a fitting with two variables determining KD (KD=0.012 ± 0.036 µM) of dimerization and kcat (0.601±0.06 min^{-1}) for activity (OriginPro) (Material and methods chapter 3.4)

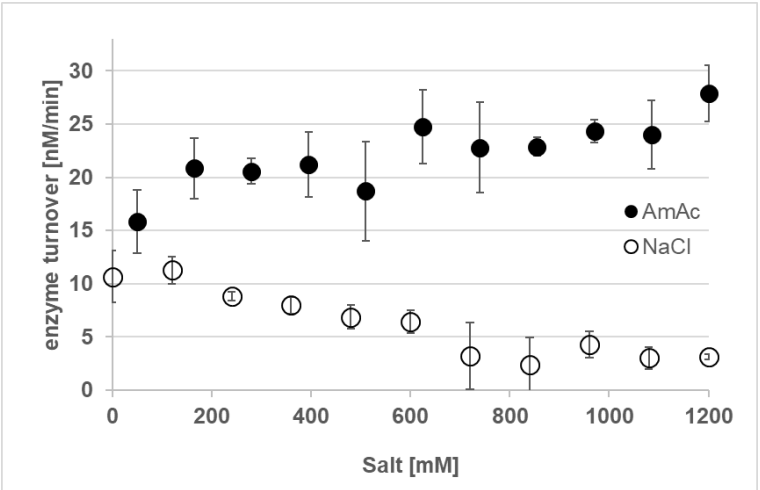


Figure S 2: Influence of AmAc and NaCl on M_{pro} enzyme efficiency. The influence of salt conditions on enzyme activity was determined by incubating 20 µL Mpro (final concentration 0.1 µM) with 80 µM FRET peptide substrate FSP4-5 (final concentration 1 µM) in varying AmAc (10% Glycerole, pH7.5 and 50 mM 165 mM 280 mM, 395 mM, 510 mM, 625 mM, 740 mM, 855 mM, 970 mM, 1085 mM and 1200 mM) and NaCl (50 mM HEPES, 10% Glycerol, pH7.5 and 0 mM, 120 mM, 240 mM, 360 mM, 480 mM, 600 mM, 720 mM, 840 mM, 960 mM, 1080 mM and 1200 mM). The fluorescence was followed every 60 seconds and the enzyme turnover was determined from the initial slopes (AFU/s). While activity of SARS Mpro is slightly increasing with AmAc concentration, it is inhibited by NaCl.

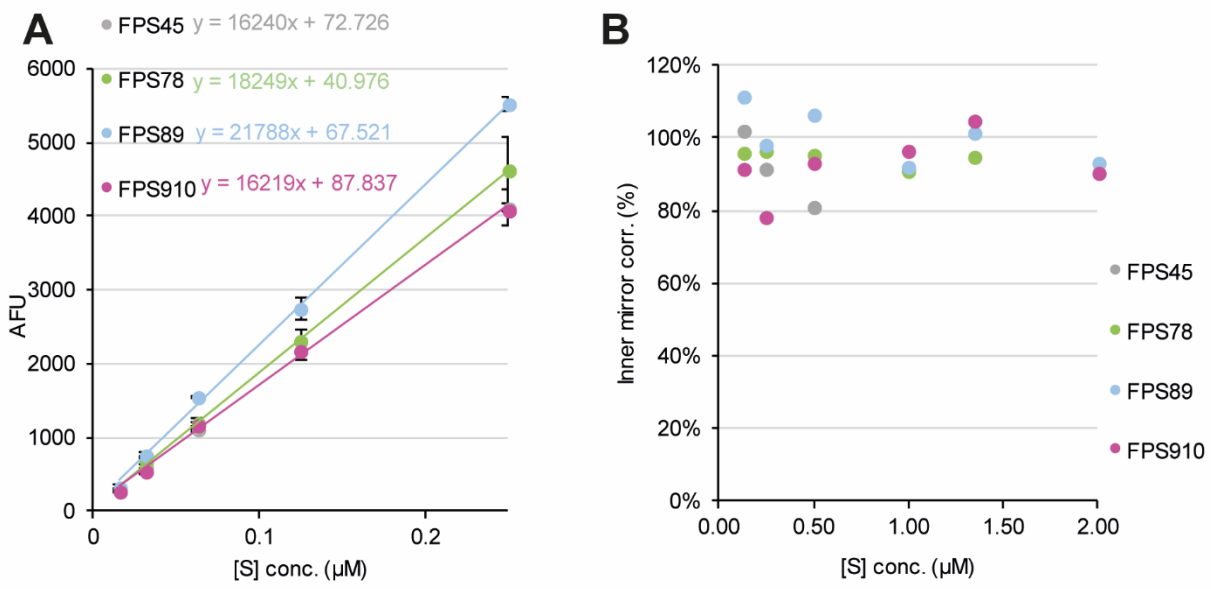


Figure S 3: Determination of specific fluorescence coefficient and inner mirror effect. **(A)** Specific FPS fluorescence determined from the slope of arbitrary fluorescence units at different concentrations of completely cleaved substrate (conc.[S]). The found specific fluorescence was used to relate measured AFU increase in activity assays to concentration of cleaved peptide. **(B)** Inner filter correction for varying substrate concentrations (conc.[S]). Since no clear trend was apparent the inner filter effect appears not to play a critical role in the used experimental setup and the correction was not applied for data analysis of other measurements.

%	SARS	TGEV	FIP	PEDV	229e
SARS	100	42.7	43.4	42.3	44.1
TGEV	42.7	100	93.9	60.8	64
FIP	43.4	93.9	100	61.5	63.3
PEDV	42.3	60.8	61.5	100	70.9
229e	44.1	64	63.3	70.9	100

Figure S 4: Sequence identity matrix of NSP7-8 from different CoV species. Analyzed with Clustal Omega Webserver. Sequences used from Table S 2: Amino acid sequences of pp1a/ab (region NSP4/5-NSP10/11/12).

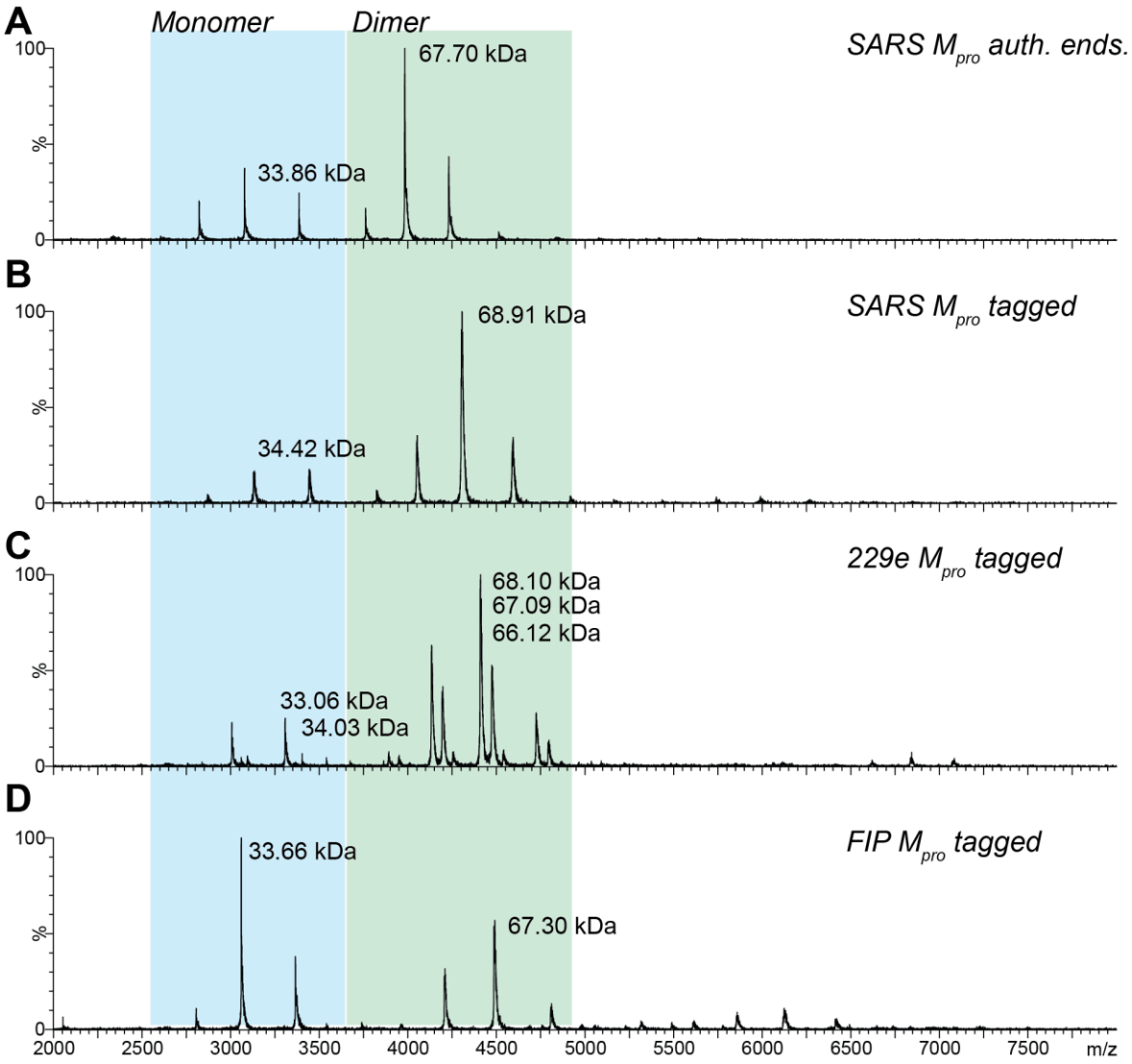


Figure S 5: Native mass spectra of different M_{pro} used during course of this work. (A) 5 µM SARS M_{pro} with authentic ends. Main protease used in experiments. (See Material and Methods 3.2 and Results of sample preparation 4.1). **(B)** 5 µM SARS M_{pro} tagged (not used for experiments shown here). **(C)** 5 µM 229e M_{pro} . Used for processing of 229e NSP7-10 (See chapter 4.2.5). Heterogeneity detected, at least two monomer masses lead triplicate size distribution in dimer. Possibly damage due to freezing or long storage. **(D)** 5 µM FIP M_{pro}. Used for cleavage of truncated PEDV NSP7-8 for complex formation (See Figure 53).

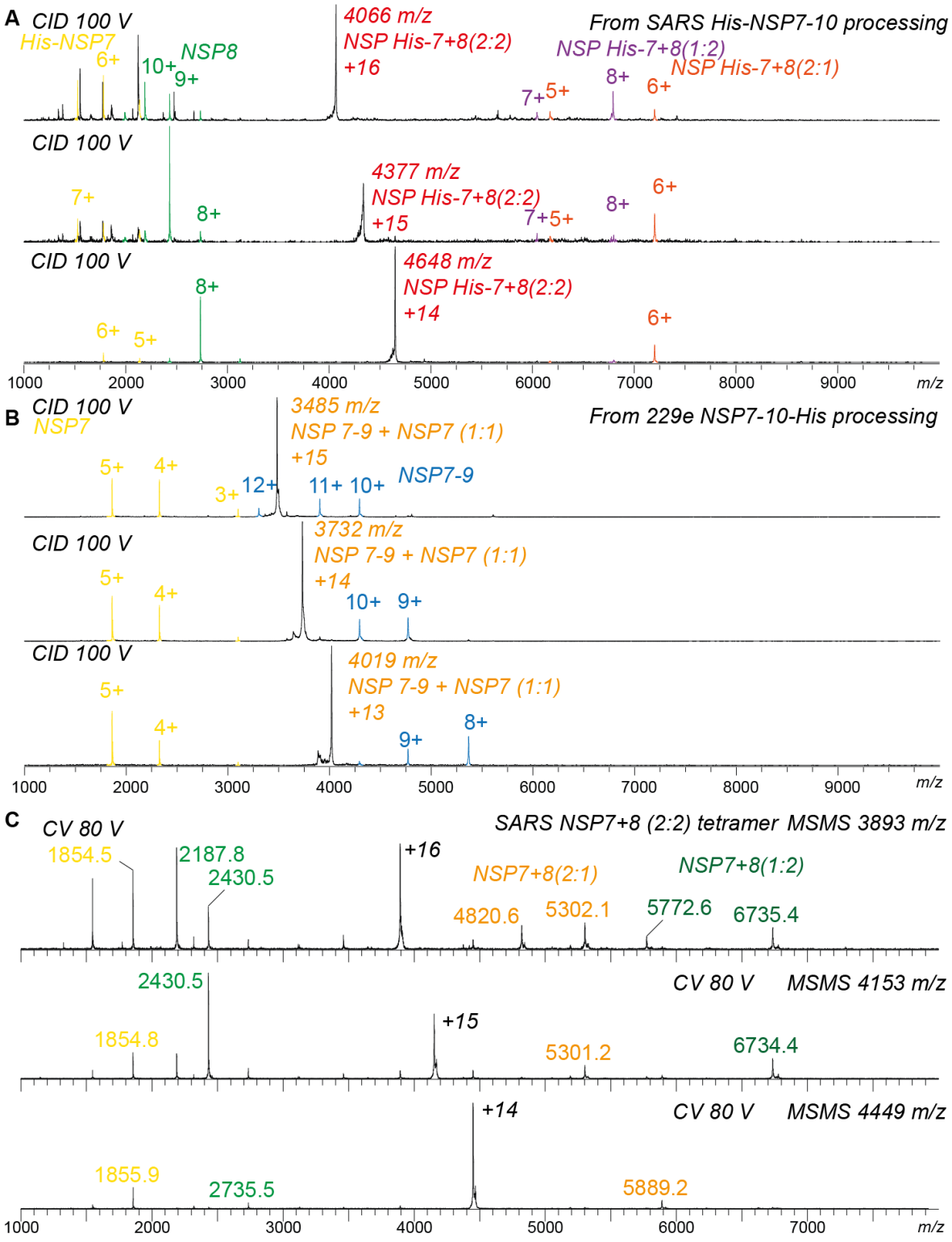


Figure S 6 CID-MS of different complexes formed upon polyprotein processing. (A) Product ion spectra of three charge states as precursors showing SARS NSP His-7+8 (2:2). Black peaks show products that could not be assigned, probably from overlapping precursor ions. **(B)** Product ion spectra of three charge states as precursors showing 229e NSP7-9 + NSP7 (1:1) (C) **(C)** Product ion spectra of three charge states as precursors showing SARS NSP 7+8 (2:2).

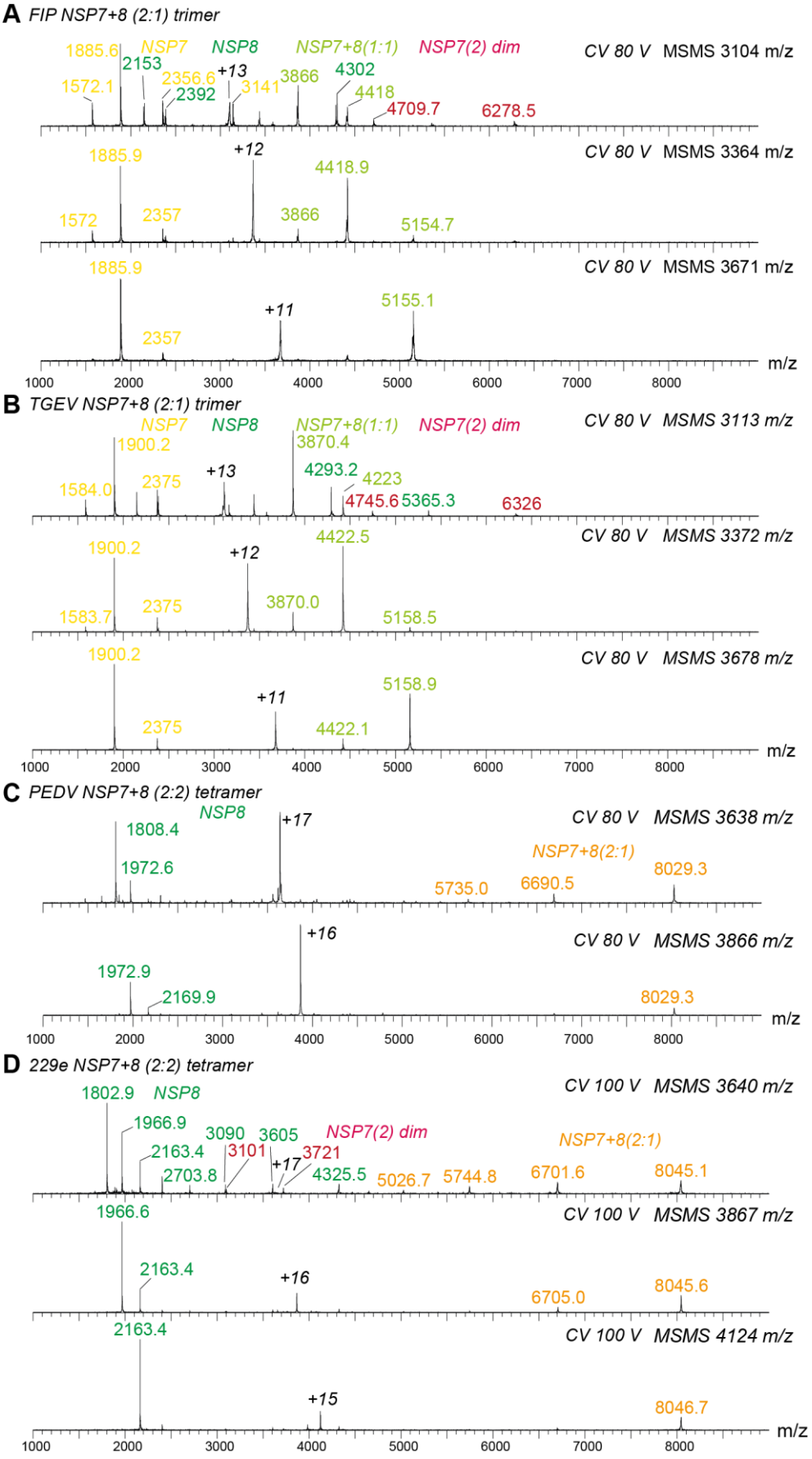


Figure S 7: CID-MS of NSP7+8 complexes of different CoV species. Product ion spectra of indicated precursor ions of **(A)** FIP **(B)** TGEV **(C)** PEDV **(D)** 229e. Product ions that reveal stoichiometry are highlighted.

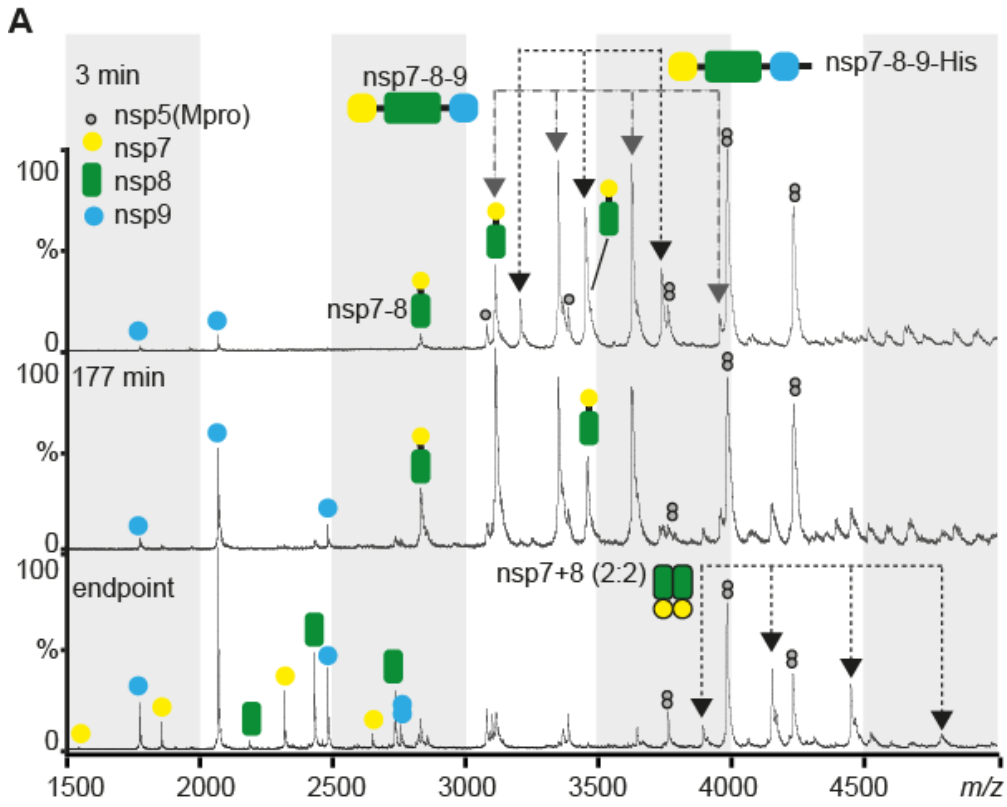


Figure S 8: Native MS exemplary spectra of SARS NSP7-9-His processing. Exemplary native MS spectra showing three time-points of sampling. For in *in vitro* processing, 2 µM M_{pro} was incubated with 14 µM NSP7-9 (ratio 1:7) at 4°C in 250 mM AmAc, 1 mM DTT, pH8. **Top:** Native mass spectrum after 3 min. **Middle:** Native mass spectrum after 177 min. **Bottom:** Native mass spectrum after 20 h.

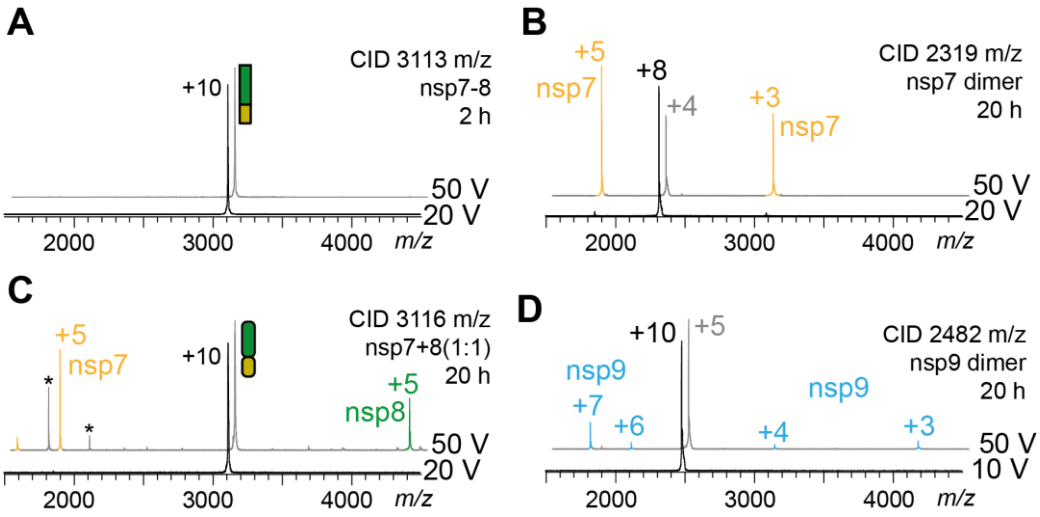


Figure S 9: CID of complexes from SARS NSP7-9 processing. Precursor ion spectra (black) and product ion spectra (grey) of putative complexes in NSP7-9 processing. Tandem MS in Q-TOF with collision cell with argon at 14 µbar. Dissociation of complexes triggered by increasing CV from 10 to 50 V. (grey and colored spectra). **(A)** CID of 10+ NSP7-8 after 2h *in vitro* processing. No dissociation of subunits confirmed the assignment of covalently bound NSP7-8.**(B)** CID of the 10+ ion after 20 h of *in vitro* processing. Dissociating subunits of His-NSP7 and NSP8 revealed formation of NSP7+8(1:1) complex. Indicated (asterisk) is an undefined mass species. **(C)** CID of NSP7 and **(D)** NSP9 homo-dimers showing subunits dissociate.

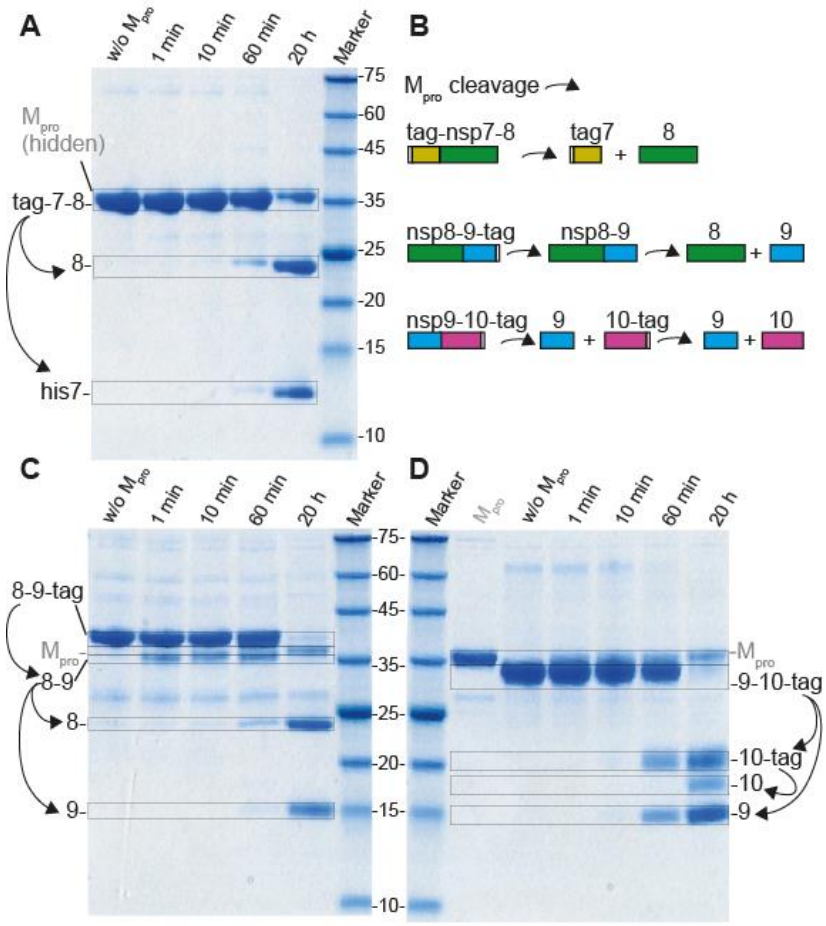


Figure S 10: SDS-PAGE: SARS NSP7-8, NSP8-9 and NSP9-10 processing. SARS NSP proteins were incubated with M_{pro} at 4°C in 20mM Phosphate buffer, 150 mM NaCl, pH8 with 1 mM DTT. Shown here is the SDS-PAGE analysis of tag-NSP7-8 (A) with an uncleavable His tag as well as NSP8-9-tag (C) and NSP9-10-tag (D), both with a cleavable His tag. After 60 minutes, NSP9-10 product bands are clearly visible, while NSP7-8 and NSP8-9 products only appear as faint bands. Between NSP7-8 and NSP9-10, there is no order of cleavage standing out. After 20 h, all cleavage sites were hydrolyzed, except NSP10-his. (B) Symbols illustrate the cleavage reactions as interpreted from the SDS-PAGE results. Arrows next to the lanes and symbols (C) illustrate these findings. SDS-PAGE was performed in 4-12% Bis-tris gel with XT MES buffer. The time points of stopping cleavage reaction are indicated in the lanes above.

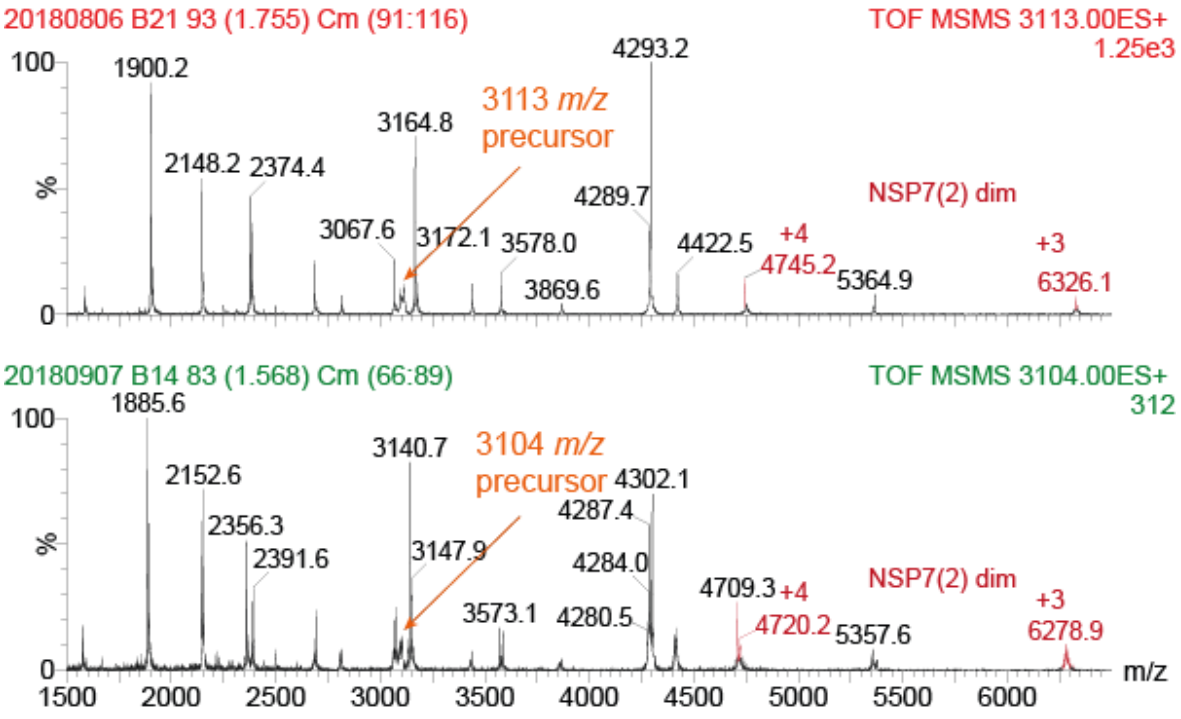


Figure S 11: Complete dissociation of precursor in CID of TGEV and FIP trimer. Product ion spectra of precursors of TGEV 3113 *m/z* and FIP 3104 *m/z*. Dissociation triggered in Q-TOF tandem MS with increasing CV at 14 µbar argon. Precursors virtually completely dissociate. NSP7(2) dimer (crimson) indicated for +3 and +4 charge state. RAW spectra taken from MassLynx 4.1 and slightly modified.

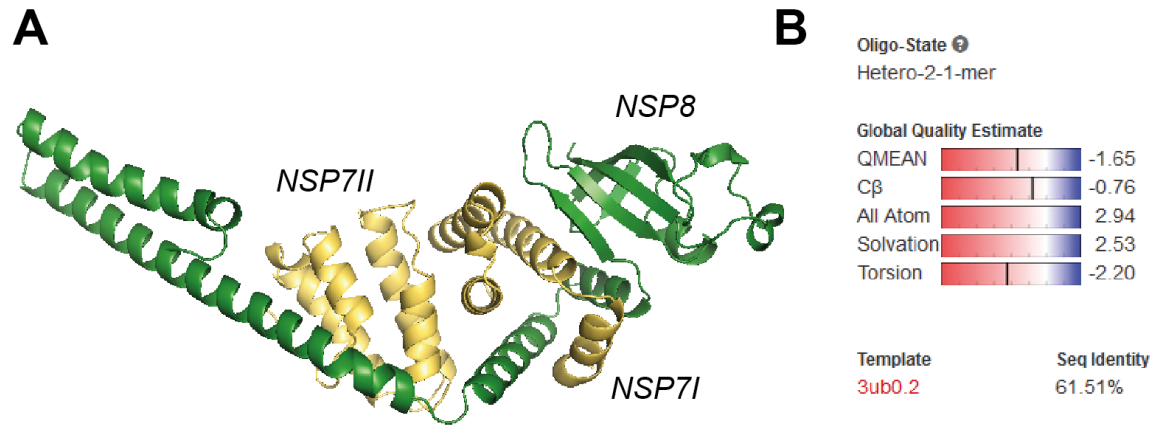


Figure S 12: Homology modelling of PEDV trimer. **(A)**Molecular structure of a PEDV NSP7+8 model generated with SWISS-MODEL, which indicated the highest *Qmean* score for homology with the hetero-trimeric NSP7+8(2:1) FIP complex. Certain properties, as the interconnection between two NSP7 molecules were confirmed by CID. Domains are indicated. **(B)** Overview of modelling parameters.

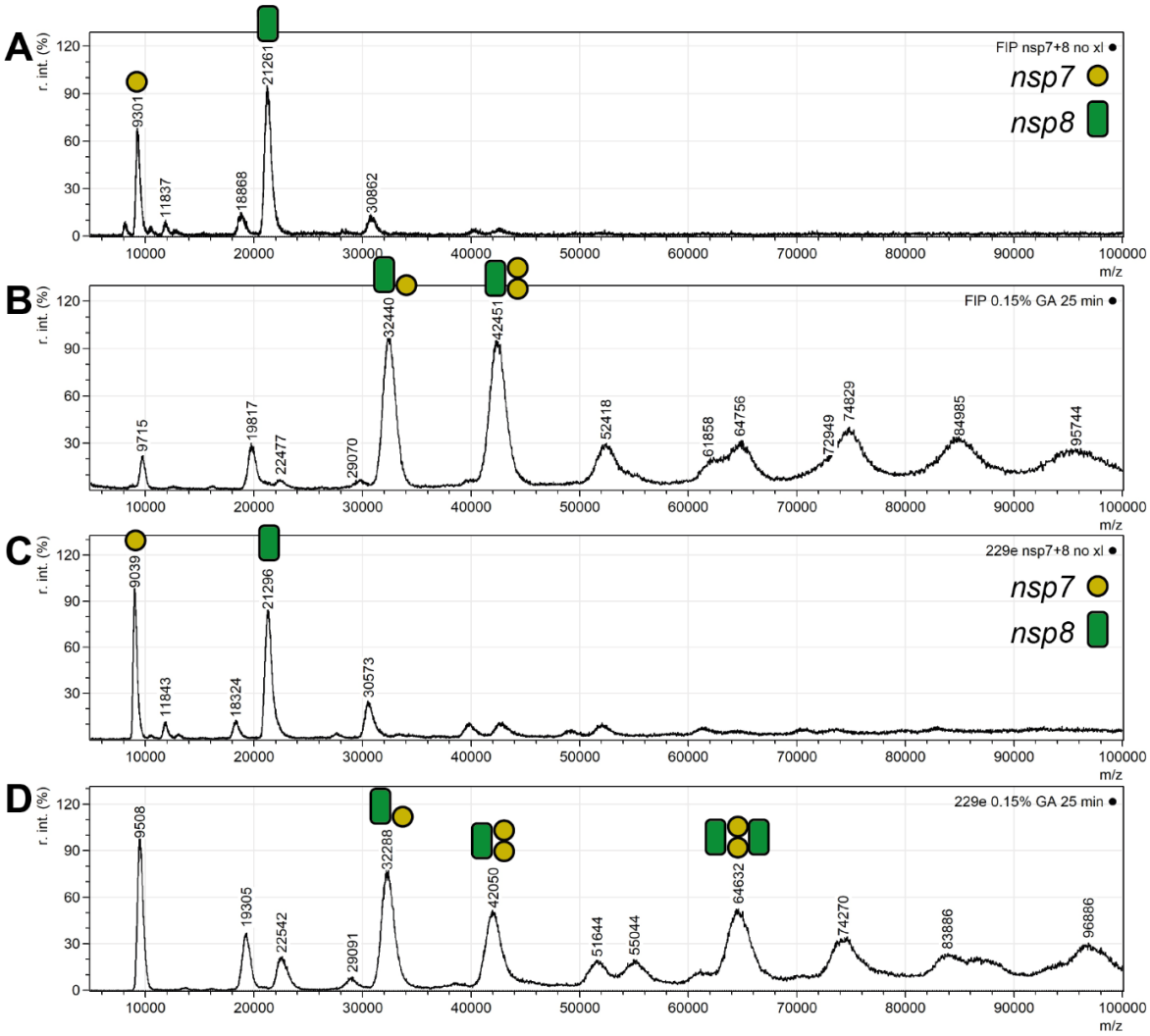


Figure S 13: MALDI-MS: Cross-linked NSP7+8 complexes of FIP and 229e. (A) MALDI mass spectra of FIP NSP7+8 showing NSP7 (9.3 kDa) and NSP8 (21.25 kDa) at their singly charged state and monomeric mass due to the disturbance of non-covalent bond in ionization. **(B)** FIP NSP7+8 0.15 % Glutaraldehyde for 25 min at 4°C. The cross-linker stabilizes the protein complexes and signals at 32.4 kDa and 42.5 kDa indicate for (1:1)dimer and (2:1) trimer. Signals above 50,000 m/z most likely due to unspecific crosslinking. (**C)** MALDI mass spectra of 229e NSP7+8 (9.5 kDa) and NSP8 (21.3 kDa). **(D)** FIP NSP7+8 0.15 % Glutaraldehyde for 25 min at 4°C. Most abundant complexes stabilized by crosslinker are indicated as (1:1) dimer (32.2 kDa), (2:1)trimer (42.1 kDa) and (2:2)tetramer (64.6 kDa). Complexes purified by size-exclusion chromatography in 20 mM HEPES, 150 mM NaCl, 4 mM ß-mercaptoethanol NSP7+8 in 20 mM HEPES, 150 mM Tris, pH8 were adjusted to 20 µM (~0.6 mg/mL). Before measurement the samples were diluted to 1 µM, immediately mixed with an acidified sinapinic acid matrix solution and then dried on a MALDI spot plate. Colored symbols depict stoichiometric mass species of NSP7 (yellow) and NSP8 (green) with highest signal strength. Deriving higher masses in cross-linked samples due to weight of the glutaraldehyde molecules. Mass spectra were not calibrated. Each spectra averaged from

triplicate measurements.

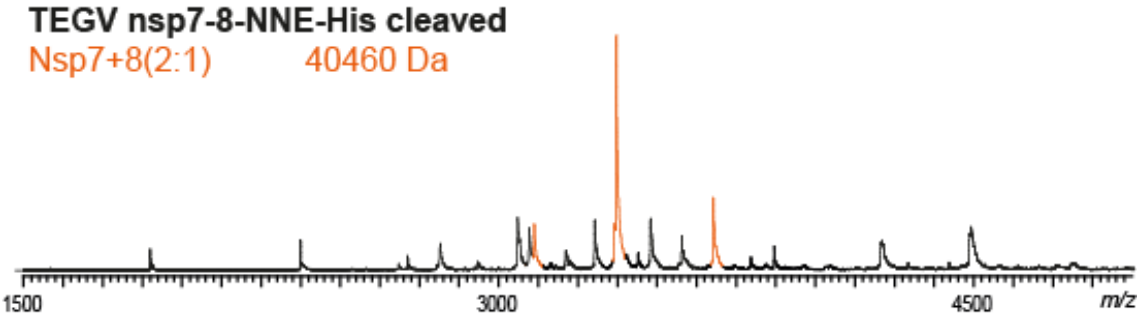


Figure S 14: TGEV trimer as most abundant species. Native mass spectrum showing NSP7+8(2:1) trimer as most intense peak series upon cleavage of TGEV NSP7-8-NNE-His. Analyzed in detail (Native MS and CID of TGEV in 4.3.2).

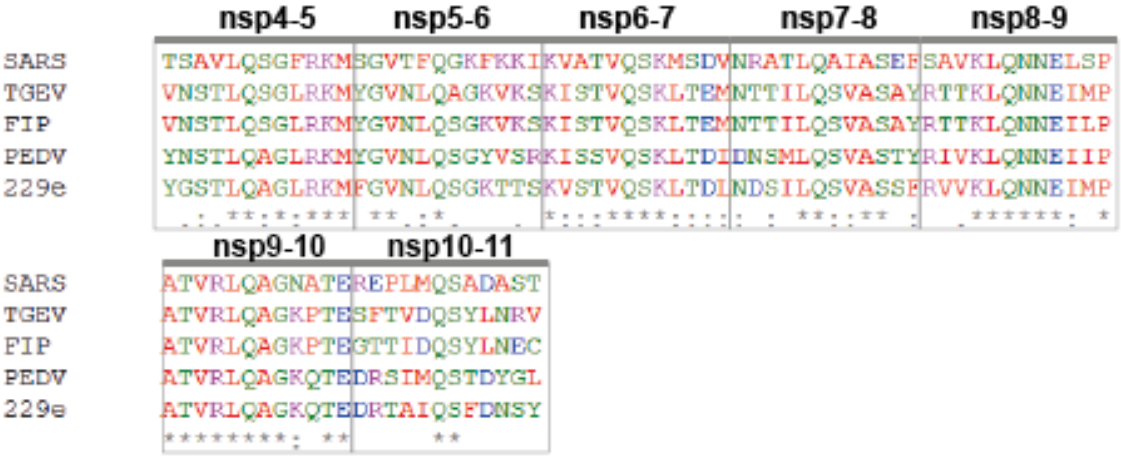


Figure S 15: Sequence alignment of cleavage sites NSP4-NSP11. Analyzed with Clustal Omega [214]. Each segment shows P1-P'6. (*)conserved residue, (:) conservative replacement, (.) semi-conservative replacement, ( ) not conserved.

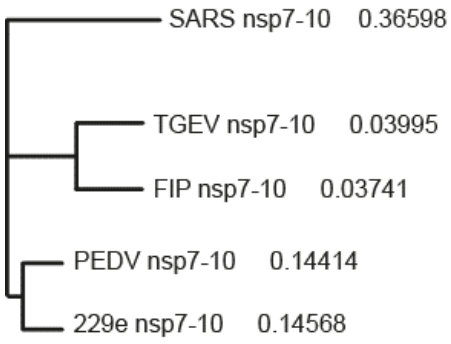


Figure S 16: Phylogenetic tree of CoV's based on NSP7-10. Analyzed with Clustal Omega [214]. Sequences used from Table S 2: Amino acid sequences of pp1a/ab (region NSP4/5-NSP10/11/12).

8 References

1. Almeida, J.D. and D. Tyrrell, *The morphology of three previously uncharacterized human respiratory viruses that grow in organ culture.* Journal of General Virology, 1967. **1**(2): p. 175-178.
2. Hamre, D. and J.J. Procknow, *A new virus isolated from the human respiratory tract.* Proceedings of the Society for Experimental Biology and Medicine, 1966. **121**(1): p. 190-193.
3. Larson, H., S.E. Reed, and D. Tyrrell, *Isolation of rhinoviruses and coronaviruses from 38 colds in adults.* Journal of medical virology, 1980. **5**(3): p. 221-229.
4. van der Hoek, L., et al., *Identification of a new human coronavirus.* Nature medicine, 2004. **10**(4): p. 368.
5. Woo, P.C., et al., *Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia.* Journal of virology, 2005. **79**(2): p. 884-895.
6. Drosten, C., et al., *Identification of a novel coronavirus in patients with severe acute respiratory syndrome.* New England Journal of Medicine, 2003. **348**(20): p. 1967-1976.
7. Wang, M., et al., *SARS-CoV infection in a restaurant from palm civet.* Emerging infectious diseases, 2005. **11**(12): p. 1860.
8. WHO, *Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003*. http://www.who.int/csr/sars/country/table2004_04_21/en/index.html, 2003.
9. Lee, J.-W. and W.J. McKibbin. *Estimating the global economic costs of SARS*. in *Learning from SARS: preparing for the next disease outbreak: workshop summary*. 2004. National Academies Press Washington, DC.
10. Chiu, W.-T., J.-S. Huang, and Y.-S. Ho, *Bibliometric analysis of severe acute respiratory syndrome-related research in the beginning stage.* Scientometrics, 2004. **61**(1): p. 69-77.
11. Wang, Z., et al., *A bibliometric analysis of PubMed literature on middle east respiratory syndrome.* International journal of environmental research and public health, 2016. **13**(6): p. 583.
12. Sweileh, W.M., *Global research trends of World Health Organization's top eight emerging pathogens.* Globalization and Health, 2017. **13**(1): p. 9.
13. Zaki, A.M., et al., *Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia.* New England Journal of Medicine, 2012. **367**(19): p. 1814-1820.
14. Sabir, J.S., et al., *Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia.* Science, 2016. **351**(6268): p. 81-84.
15. *WHO Disease Outbreak News April, 26th of 2016*. https://www.who.int/csr/don/26-april-2016-mers-saudi-arabia/en/.
16. *WHO MERS Update February 2019*. http://www.emro.who.int/pandemic-epidemic-diseases/mers-cov/mers-situation-update-february-2019.html.
17. Menachery, V.D., R.L. Graham, and R.S. Baric, *Jumping species-a mechanism for coronavirus persistence and survival.* Curr Opin Virol, 2017. **23**: p. 1-7.
18. Li, W., et al., *Bats are natural reservoirs of SARS-like coronaviruses.* Science, 2005. **310**(5748): p. 676-679.
19. Woo, P.C.Y., et al., *Discovery of Seven Novel Mammalian and Avian Coronaviruses* Journal of Virology, 2012. **86**(7): p. 3995-4008.
20. Wood, E., *An apparently new syndrome of porcine epidemic diarrhoea.* The Veterinary record, 1977. **100**(12): p. 243.
21. Bevins, S.N., et al., *Spillover of swine coronaviruses, United States.* Emerging infectious diseases, 2018. **24**(7): p. 1390.
22. Holzworth, J., *Some important disorders of cats.* The Cornell Veterinarian, 1963. **53**: p. 157-160.
23. Vennema, H., et al., *Feline infectious peritonitis viruses arise by mutation from endemic feline enteric coronaviruses.* Virology, 1998. **243**(1): p. 150-157.
24. Ge, X.-Y., et al., *Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor.* Nature, 2013. **503**(7477): p. 535.
25. Walls, A.C., et al., *Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer.* Nature, 2016. **531**(7592): p. 114-7.
26. Walls, A.C., et al., *Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion.* Proceedings of the National Academy of Sciences, 2017. **114**(42): p. 11157.
27. Walls, A.C., et al., *Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion.* Cell, 2019.
28. Bosch, B.J., et al., *The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex.* Journal of virology, 2003. **77**(16): p. 8801-8811.
29. Krijnse-Locker, J., et al., *Characterization of the budding compartment of mouse hepatitis virus: evidence that transport from the RER to the Golgi complex requires only one vesicular transport step.* The Journal of Cell Biology, 1994. **124**(1): p. 55-70.

30. Perlman, S. and J. Netland, *Coronaviruses post-SARS: update on replication and pathogenesis.* Nat Rev Microbiol, 2009. **7**(6): p. 439-50.
31. Fehr, A.R. and S. Perlman, *Coronaviruses: an overview of their replication and pathogenesis*, in *Coronaviruses*. 2015, Springer. p. 1-23.
32. Lefkowitz, E.J., et al., *Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV).* Nucleic acids research, 2017. **46**(D1): p. D708-D717.
33. Desper, R. and O. Gascuel, *Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle.* J Comput Biol, 2002. **9**(5): p. 687-705.
34. Brierley, I., P. Digard, and S.C. Inglis, *Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot.* Cell, 1989. **57**(4): p. 537-47.
35. Prentice, E., et al., *Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins.* Journal of virology, 2004. **78**(18): p. 9977-9986.
36. Hsu, M.-F., et al., *Mechanism of the maturation process of SARS-CoV 3CL protease.* Journal of Biological Chemistry, 2005. **280**(35): p. 31257-31266.
37. Marra, M.A., et al., *The Genome sequence of the SARS-associated coronavirus.* Science, 2003. **300**.
38. Thiel, V., et al., *Mechanisms and enzymes involved in SARS coronavirus genome expression.* Journal of General Virology, 2003. **84**(9): p. 2305-2315.
39. Snijder, E.J., et al., *Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage.* J Mol Biol, 2003. **331**(5): p. 991-1004.
40. Gorbalenya, A.E., et al., *Nidovirales: evolving the largest RNA virus genome.* Virus research, 2006. **117**(1): p. 17-37.
41. Oudshoorn, D., et al., *Expression and cleavage of Middle East respiratory syndrome coronavirus nsp3-4 polyprotein induce the formation of double-membrane vesicles that mimic those associated with coronaviral RNA replication.* mBio, 2017. **8**(6): p. e01658-17.
42. Romero-Brey, I. and R. Bartenschlager, *Endoplasmic reticulum: the favorite intracellular niche for viral replication and assembly.* Viruses, 2016. **8**(6): p. 160.
43. Knoops, K., et al., *SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum.* PLoS biology, 2008. **6**(9): p. e226.
44. V'kovski, P., et al., *Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling.* eLife, 2019. **8**: p. e42037.
45. Hagemeijer, M.C., et al., *Dynamics of coronavirus replication-transcription complexes.* Journal of virology, 2010. **84**(4): p. 2134-2149.
46. Van Hemert, M.J., et al., *SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro.* PLoS pathogens, 2008. **4**(5): p. e1000054.
47. Gorbalenya, A.E., et al., *Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis.* Nucleic Acids Research, 1989. **17**(12): p. 4847-4861.
48. de Wilde, A.H., et al., *Host Factors in Coronavirus Replication.* 2017.
49. Anand, K., et al., *Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs.* Science, 2003. **300**(5626): p. 1763-1767.
50. Kuo, C.-J., et al., *Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate.* Biochemical and biophysical research communications, 2004. **318**(4): p. 862-867.
51. Anand, K., et al., *Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain.* The EMBO Journal, 2002. **21**(13): p. 3213-3224.
52. Xia, B. and X. Kang, *Activation and maturation of SARS-CoV main protease.* Protein & cell, 2011. **2**(4): p. 282-290.
53. Chen, S., et al., *Liberation of SARS-CoV main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode.* Protein Cell, 2010. **1**(1): p. 59-74.
54. Li, C., et al., *Maturation mechanism of severe acute respiratory syndrome (SARS) coronavirus 3C-like proteinase.* J Biol Chem, 2010. **285**(36): p. 28134-40.
55. Schiller, J.J., A. Kanjanahaluethai, and S.C. Baker, *Processing of the coronavirus MHV-JHM polymerase polyprotein: identification of precursors and proteolytic products spanning 400 kilodaltons of ORF1a.* Virology, 1998. **242**(2): p. 288-302.
56. Xue, X., et al., *Production of authentic SARS-CoV M pro with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction.* Journal of molecular biology, 2007. **366**(3): p. 965-975.
57. St. John, S.E., B.J. Anson, and A.D. Mesecar, *X-Ray Structure and Inhibition of 3C-like Protease from Porcine Epidemic Diarrhea Virus.* Scientific Reports, 2016. **6**: p. 25961.
58. Yang, H., et al., *The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor.* Proceedings of the National Academy of Sciences, 2003. **100**(23): p. 13190-13195.

59. Zhang, S., et al., *Three-dimensional domain swapping as a mechanism to lock the active conformation in a super-active octamer of SARS-CoV main protease.* Protein & cell, 2010. **1**(4): p. 371-383.
60. Fan, K., et al., *Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase.* J Biol Chem, 2004. **279**(3): p. 1637-42.
61. Chen, H., et al., *Only one protomer is active in the dimer of SARS 3C-like proteinase.* J Biol Chem, 2006. **281**(20): p. 13894-8.
62. Shi, J., J. Sivaraman, and J. Song, *Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease.* Journal of virology, 2008. **82**(9): p. 4620-4629.
63. Tomar, S., et al., *Ligand-induced Dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus nsp5 Protease (3CL(pro)): IMPLICATIONS FOR nsp5 REGULATION AND THE DEVELOPMENT OF ANTIVIRALS.* The Journal of Biological Chemistry, 2015. **290**(32): p. 19403-19422.
64. Grum-Tokars, V., et al., *Evaluating the 3C-like protease activity of SARS-Coronavirus: recommendations for standardized assays for drug discovery.* Virus research, 2008. **133**(1): p. 63-73.
65. Chen, S., et al., *Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase N Terminus Is Indispensable for Proteolytic Activity but Not for Enzyme Dimerization BIOCHEMICAL AND THERMODYNAMIC INVESTIGATION IN CONJUNCTION WITH MOLECULAR DYNAMICS SIMULATIONS.* Journal of Biological Chemistry, 2005. **280**(1): p. 164-173.
66. Hsu, W.-C., et al., *Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease.* Journal of Biological Chemistry, 2005. **280**(24): p. 22741-22748.
67. WEI, P., et al., *Substrate Binding and Homo-Dimerization of SARS 3CL Proteinase are Mutual Allosteric Effectors.* Acta Physico-Chimica Sinica, 2010. **26**(4): p. 1093-1098.
68. Cheng, S.-C., G.-G. Chang, and C.-Y. Chou, *Mutation of Glu-166 blocks the substrate-induced dimerization of SARS coronavirus main protease.* Biophysical journal, 2010. **98**(7): p. 1327-1336.
69. Chu, L.H.M., et al., *Rapid peptide-based screening on the substrate specificity of severe acute respiratory syndrome (SARS) coronavirus 3C-like protease by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.* Protein science, 2006. **15**(4): p. 699-709.
70. Goetz, D., et al., *Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the SARS coronavirus.* Biochemistry, 2007. **46**(30): p. 8744-8752.
71. Chuck, C.-P., et al., *Profiling of substrate specificity of SARS-CoV 3CLpro.* PloS one, 2010. **5**(10): p. e13197.
72. Chuck, C.-P., et al., *Profiling of substrate specificities of 3C-like proteases from group 1, 2a, 2b, and 3 coronaviruses.* PloS one, 2011. **6**(11): p. e27228.
73. Oostra, M., et al., *Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning.* Journal of virology, 2008. **82**(24): p. 12392-12405.
74. Fan, K., et al., *Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase.* Journal of Biological Chemistry, 2004. **279**(3): p. 1637-1642.
75. Wu, A., et al., *Prediction and biochemical analysis of putative cleavage sites of the 3C-like protease of Middle East respiratory syndrome coronavirus.* Virus Research, 2015. **208**: p. 56-65.
76. Sawicki, S.G., et al., *Functional and genetic analysis of coronavirus replicase-transcriptase proteins.* PLoS Pathog, 2005. **1**(4): p. e39.
77. Tan, Y.W., et al., *Coronavirus infectious bronchitis virus non-structural proteins 8 and 12 form stable complex independent of the non-translated regions of viral RNA and other viral proteins.* Virology, 2018. **513**: p. 75-84.
78. Deming, D.J., et al., *Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication.* Journal of virology, 2007. **81**(19): p. 10280-10291.
79. Lu, X.T., A.C. Sims, and M.R. Denison, *Mouse hepatitis virus 3C-like protease cleaves a 22-kilodalton protein from the open reading frame 1a polyprotein in virus-infected cells and in vitro.* Journal of virology, 1998. **72**(3): p. 2265-2271.
80. Liu, D., H. Xu, and T. Brown, *Proteolytic processing of the coronavirus infectious bronchitis virus 1a polyprotein: identification of a 10-kilodalton polypeptide and determination of its cleavage sites.* Journal of virology, 1997. **71**(3): p. 1814-1820.
81. Ziebuhr, J. and S.G. Siddell, *Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: Identification of proteolytic products and cleavage sites common to pp1a and pp1ab.* Journal of Virology, 1999. **73**(1): p. 177-185.
82. Hegyi, A. and J. Ziebuhr, *Conservation of substrate specificities among coronavirus main proteases.* J Gen Virol, 2002. **83**(Pt 3): p. 595-9.
83. Falke, S., *Coronaviral Polyprotein Nsp7-10 : Proteolytic Processing and Dynamic Interactions within the Transcriptase/Replicase Complex*. 2014, Staats- und Universitätsbibliothek Hamburg: Hamburg.

84. Xiao, Y.B., et al., *Nonstructural Proteins 7 and 8 of Feline Coronavirus Form a 2:1 Heterotrimer That Exhibits Primer-Independent RNA Polymerase Activity.* Journal of Virology, 2012. **86**(8): p. 4444-4454.
85. Peti, W., et al., *Structural genomics of the severe acute respiratory syndrome coronavirus: nuclear magnetic resonance structure of the protein nsP7.* Journal of virology, 2005. **79**(20): p. 12905-12913.
86. Zhai, Y.J., et al., *Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer.* Nature Structural & Molecular Biology, 2005. **12**(11): p. 980-986.
87. Li, S., et al., *New nsp8 isoform suggests mechanism for tuning viral RNA synthesis.* Protein Cell, 2010. **1**(2): p. 198-204.
88. Te Velthuis, A.J., et al., *The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent.* Nucleic acids research, 2010. **38**(1): p. 203-214.
89. te Velthuis, A.J., S.H. van den Worm, and E.J. Snijder, *The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension.* Nucleic Acids Res, 2012. **40**(4): p. 1737-47.
90. Imbert, I., et al., *A second, non-canonical RNA-dependent RNA polymerase in SARS Coronavirus.* The EMBO journal, 2006. **25**(20): p. 4933-4942.
91. Subissi, L., et al., *One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities.* Proceedings of the National Academy of Sciences, 2014. **111**(37): p. E3900-E3909.
92. Tvarogová, J., et al., *Identification and characterization of a human coronavirus 229E nonstructural protein 8-associated RNA 3′ -terminal adenylyltransferase activity.* Journal of virology, 2019: p. JVI. 00291-19.
93. Kirchdoerfer, R.N. and A.B. Ward, *Structure of the SARS-CoV NSP12 polymerase bound to NSP7 and NSP8 co-factors.* bioRxiv, 2019: p. 551986.
94. Su, D., et al., *Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10.* Journal of virology, 2006. **80**(16): p. 7902-7908.
95. Matthes, N., et al., *The non-structural protein Nsp10 of mouse hepatitis virus binds zinc ions and nucleic acids.* FEBS letters, 2006. **580**(17): p. 4143-4149.
96. Joseph, J.S., et al., *Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs.* Journal of virology, 2006. **80**(16): p. 7894-7901.
97. Bouvet, M., et al., *RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex.* Proceedings of the National Academy of Sciences, 2012. **109**(24): p. 9372-9377.
98. Decroly, E., et al., *Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2′ -O-methyltransferase nsp10/nsp16 complex.* PLoS pathogens, 2011. **7**(5): p. e1002059.
99. Ferron, F., et al., *Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA.* Proceedings of the National Academy of Sciences, 2018. **115**(2): p. E162-E171.
100. Ma, Y., et al., *Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex.* Proc Natl Acad Sci U S A, 2015.
101. Minskaia, E., et al., *Discovery of an RNA virus 3′ → 5′ exoribonuclease that is critically involved in coronavirus RNA synthesis.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(13): p. 5108-5113.
102. Eckerle, L.D., et al., *High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants.* Journal of virology, 2007. **81**(22): p. 12135-12144.
103. Eckerle, L.D., et al., *Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing.* PLoS pathogens, 2010. **6**(5): p. e1000896.
104. Lugari, A., et al., *Molecular mapping of the RNA Cap 2′ -O-methyltransferase activation interface between severe acute respiratory syndrome coronavirus nsp10 and nsp16.* Journal of Biological Chemistry, 2010. **285**(43): p. 33230-33241.
105. Bouvet, M., et al., *In vitro reconstitution of SARS-coronavirus mRNA cap methylation.* PLoS pathogens, 2010. **6**(4): p. e1000863.
106. Martin, J.L. and F.M. McMillan, *SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold.* Current opinion in structural biology, 2002. **12**(6): p. 783-793.
107. Miknis, Z.J., et al., *Severe Acute Respiratory Syndrome Coronavirus nsp9 Dimerization Is Essential for Efficient Viral Growth.* Journal of Virology, 2008. **83**(7): p. 3007–3018.
108. Egloff, M.P., et al., *The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world.* Proc Natl Acad Sci U S A, 2004. **101**(11): p. 3792-6.
109. Sutton, G., et al., *The nsp9 Replicase Protein of SARS-Coronavirus, Structure and Functional Insights.* Structure, 2004. **12**(2).
110. Zeng, Z., et al., *Dimerization of Coronavirus nsp9 with Diverse Modes Enhances Its Nucleic Acid Binding Affinity.* Journal of Virology, 2018. **92**(17): p. e00692-18.

111. Miknis, Z.J., et al., *Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth.* Journal of virology, 2009. **83**(7): p. 3007-3018.
112. Tingting, H., et al., *Structural basis for dimerization and RNA binding of avian infectious bronchitis virus nsp9.* Protein Science, 2017. **26**(5).
113. Egloff, M.-P., et al., *The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world.* Proceedings of the National Academy of Sciences of the United States, 2004. **101**(11): p. 3792–3796.
114. Ponnusamy, R., et al., *Variable oligomerization modes in coronavirus non-structural protein 9.* J Mol Biol, 2008. **383**(5): p. 1081-96.
115. Hu, T., et al., *Structural basis for dimerization and RNA binding of avian infectious bronchitis virus nsp9.* Protein Science, 2017. **26**(5): p. 1037-1048.
116. Baez-Santos, Y.M., S.E. St. John, and A.D. Mesecar, *The SARS-Coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds.* Antiviral Research, 2015. **115**: p. 21-38.
117. Dülfer, K., Kopicki, Krichel, Uetrecht, *Structural mass spectrometry goes viral.* 2019.
118. Heidemann, J., B. Krichel, and C. Uetrecht, *Native Massenspektrometrie für die Proteinstrukturanalytik.* BIOspektrum, 2018. **24**(2): p. 164-167.
119. Merk, A., et al., *Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery.* Cell, 2016. **165**(7): p. 1698-1707.
120. Ward, A.B., A. Sali, and I.A. Wilson, *Biochemistry. Integrative structural biology.* Science, 2013. **339**(6122): p. 913-5.
121. Politis, A. and C. Schmidt, *Structural characterisation of medically relevant protein assemblies by integrating mass spectrometry with computational modelling.* J Proteomics, 2018. **175**: p. 34-41.
122. Marklund, E.G. and J.L. Benesch, *Weighing-up protein dynamics: the combination of native mass spectrometry and molecular dynamics simulations.* Curr Opin Struct Biol, 2019. **54**: p. 50-58.
123. Yamashita, M. and J.B. Fenn, *Electrospray Ion-Source - Another Variation on the Free-Jet Theme.* Journal of Physical Chemistry, 1984. **88**(20): p. 4451-4459.
124. Karas, M., et al., *Matrix-assisted ultraviolet laser desorption of non-volatile compounds.* International journal of mass spectrometry and ion processes, 1987. **78**: p. 53-68.
125. Tanaka, K., et al., *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry.* Rapid communications in mass spectrometry, 1988. **2**(8): p. 151-153.
126. Loo, J.A., *Observation of Large Subunit Protein Complexes by Electrospray-Ionization Mass-Spectrometry.* Journal of Mass Spectrometry, 1995. **30**(1): p. 180-183.
127. Rostom, A.A. and C.V. Robinson, *Detection of the intact GroEL chaperonin assembly by mass spectrometry.* Journal of the American Chemical Society, 1999. **121**(19): p. 4718-4719.
128. Van Berkel, W.J.H., et al., *Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry.* Protein Science, 2000. **9**(3): p. 435-439.
129. Leney, A.C. and A.J.R. Heck, *Native Mass Spectrometry: What is in the Name?* Journal of the American Society for Mass Spectrometry, 2017. **28**(1): p. 5-13.
130. Katta, V. and B.T. Chait, *Conformational-Changes in Proteins Probed by Hydrogen-Exchange Electrospray-Ionization Mass-Spectrometry.* Rapid Communications in Mass Spectrometry, 1991. **5**(4): p. 214-217.
131. Rossi, V., et al., *Structure of the Catalytic Region of Human-Complement Protease C(1)over-Bar-S - Study by Chemical Cross-Linking and 3-Dimensional Homology Modeling.* Biochemistry, 1995. **34**(22): p. 7311-7321.
132. Fenn, J.B., et al., *Electrospray Ionization for Mass-Spectrometry of Large Biomolecules.* Science, 1989. **246**(4926): p. 64-71.
133. Leney, A.C. and A.J. Heck, *Native Mass Spectrometry: What is in the Name?* J Am Soc Mass Spectrom, 2016.
134. Peetz, O., et al., *LILBID and nESI: Different Native Mass Spectrometry Techniques as Tools in Structural Biology.* Journal of The American Society for Mass Spectrometry, 2019. **30**(1): p. 181-191.
135. Konermann, L., *Addressing a Common Misconception: Ammonium Acetate as Neutral pH "Buffer" for Native Electrospray Mass Spectrometry.* Journal of The American Society for Mass Spectrometry, 2017. **28**(9): p. 1827-1835.
136. Uetrecht, C. and A.J. Heck, *Modern biomolecular mass spectrometry and its role in studying virus structure, dynamics, and assembly.* Angew Chem Int Ed Engl, 2011. **50**(36): p. 8248-62.
137. Uetrecht, C., et al., *High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids.* Proceedings of the National Academy of Sciences, 2008. **105**(27): p. 9216-9220.
138. Hogan, C.J., Jr., et al., *Charge reduced electrospray size spectrometry of mega- and gigadalton complexes: whole viruses and virus fragments.* Anal Chem, 2006. **78**(3): p. 844-52.
139. Siuzdak, G., et al., *Mass spectrometry and viral analysis.* Chem Biol, 1996. **3**(1): p. 45-8.

140. Wilm, M.S. and M. Mann, *Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?* International Journal of Mass Spectrometry and Ion Processes, 1994. **136**(2): p. 167-180.
141. McAllister, R.G., et al., *Release of native-like gaseous proteins from electrospray droplets via the charged residue mechanism: insights from molecular dynamics simulations.* Journal of the American Chemical Society, 2015. **137**(39): p. 12667-12676.
142. Konermann, L., et al., *Unraveling the Mechanism of Electrospray Ionization.* Analytical Chemistry, 2013. **85**(1): p. 2-9.
143. Kebarle, P. and U.H. Verkerk, *Electrospray: from ions in solution to ions in the gas phase, what we know now.* Mass spectrometry reviews, 2009. **28**(6): p. 898-917.
144. Rayleigh, L., *XX. On the equilibrium of liquid conducting masses charged with electricity.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1882. **14**(87): p. 184-186.
145. Smith, J.N., R.C. Flagan, and J. Beauchamp, *Droplet evaporation and discharge dynamics in electrospray ionization.* The Journal of Physical Chemistry A, 2002. **106**(42): p. 9957-9967.
146. Heck, A.J. and R.H. Van Den Heuvel, *Investigation of intact protein complexes by mass spectrometry.* Mass Spectrom Rev, 2004. **23**(5): p. 368-89.
147. Loscertales, I. and J. Fernández De La Mora, *Experiments on the kinetics of field evaporation of small ions from droplets.* The Journal of chemical physics, 1995. **103**(12): p. 5041-5060.
148. Metwally, H., Q. Duez, and L. Konermann, *Chain Ejection Model for Electrospray Ionization of Unfolded Proteins: Evidence from Atomistic Simulations and Ion Mobility Spectrometry.* Analytical Chemistry, 2018. **90**(16): p. 10069-10077.
149. Covey, T.R., et al., *The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry.* Rapid Commun Mass Spectrom, 1988. **2**(11): p. 249-56.
150. Brini, E., et al., *How water's properties are encoded in its molecular structure and energies.* Chemical reviews, 2017. **117**(19): p. 12385-12414.
151. Breuker, K. and F.W. McLafferty, *Stepwise evolution of protein native structure with electrospray into the gas phase, 10− 12 to 102 s.* Proceedings of the National Academy of Sciences, 2008. **105**(47): p. 18145-18152.
152. Bakhtiari, M. and L. Konermann, *Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures.* The Journal of Physical Chemistry B, 2019. **123**(8): p. 1784-1796.
153. Clemmer, D.E., D.H. Russell, and E.R. Williams, *Characterizing the Conformationome: Toward a Structural Understanding of the Proteome.* Accounts of Chemical Research, 2017. **50**(3): p. 556-560.
154. Fernandez De La Mora, J., *Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism.* Analytica chimica acta, 2000. **406**(1): p. 93-104.
155. Kaltashov, I.A. and A. Mohimen, *Estimates of protein surface areas in solution by electrospray ionization mass spectrometry.* Anal Chem, 2005. **77**(16): p. 5370-9.
156. Heck, A.J. and R.H. van den Heuvel, *Investigation of intact protein complexes by mass spectrometry.* Mass spectrometry reviews, 2004. **23**(5): p. 368-389.
157. Wyttenbach, T. and M.T. Bowers, *Structural stability from solution to the gas phase: native solution structure of ubiquitin survives analysis in a solvent-free ion mobility–mass spectrometry environment.* The journal of physical chemistry B, 2011. **115**(42): p. 12266-12275.
158. Scarff, C.A., et al., *Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements.* Rapid Commun Mass Spectrom, 2008. **22**(20): p. 3297-304.
159. Seo, J., et al., *Retention of Native Protein Structures in the Absence of Solvent: A Coupled Ion Mobility and Spectroscopic Study.* Angew Chem Int Ed Engl, 2016. **55**(45): p. 14173-14176.
160. Silveira, J.A., et al., *From solution to the gas phase: stepwise dehydration and kinetic trapping of substance P reveals the origin of peptide conformations.* J Am Chem Soc, 2013. **135**(51): p. 19147-53.
161. Fort, K.L., et al., *Expanding the structural analysis capabilities on an Orbitrap-based mass spectrometer for large macromolecular complexes.* Analyst, 2017. **143**(1): p. 100-105.
162. Chernushevich, I.V., A.V. Loboda, and B.A. Thomson, *An introduction to quadrupole–time-of-flight mass spectrometry.* J Mass Spectrom, 2001. **36**(8): p. 849-865.
163. Mamyrin, B., et al., *The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution.* Zh. Eksp. Teor. Fiz, 1973. **64**: p. 82-89.
164. Koppenaal, D.W., et al., *MS detectors*. 2005, ACS Publications.
165. van den Heuvel, R.H., et al., *Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry.* Anal Chem, 2006. **78**(21): p. 7473-83.
166. Paul, W., *Electromagnetic traps for charged and neutral particles (Nobel lecture).* Angewandte Chemie International Edition, 1990. **29**(7): p. 739-748.
167. Pagel, K., et al., *Alternate dissociation pathways identified in charge-reduced protein complex ions.* Anal Chem, 2010. **82**(12): p. 5363-72.

168. Boeri Erba, E. and C. Petosa, *The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes.* Protein Sci, 2015. **24**(8): p. 1176-92.
169. Gabelica, V., et al., *Influence of response factors on determining equilibrium association constants of non-covalent complexes by electrospray ionization mass spectrometry.* Journal of mass spectrometry, 2003. **38**(5): p. 491-501.
170. Root, K., et al., *Insight into Signal Response of Protein Ions in Native ESI-MS from the Analysis of Model Mixtures of Covalently Linked Protein Oligomers.* J Am Soc Mass Spectrom, 2017.
171. Wang, G., A.J. Johnson, and I.A. Kaltashov, *Evaluation of electrospray ionization mass spectrometry as a tool for characterization of small soluble protein aggregates.* Analytical chemistry, 2012. **84**(3): p. 1718-1724.
172. Wang, G., et al., *Molecular basis of assembly and activation of complement component C1 in complex with immunoglobulin G1 and antigen.* Molecular cell, 2016. **63**(1): p. 135-145.
173. Konermann, L., J. Pan, and Y.H. Liu, *Hydrogen exchange mass spectrometry for studying protein structure and dynamics.* Chem Soc Rev, 2011. **40**(3): p. 1224-34.
174. Trabjerg, E., Z.E. Nazari, and K.D. Rand, *Conformational Analysis of Complex Protein States by Hydrogen/Deuterium Exchange Mass Spectrometry (HDX-MS): Challenges and Emerging Solutions.* TrAC Trends in Analytical Chemistry, 2018.
175. Englander, S.W., *Hydrogen exchange and mass spectrometry: A historical perspective.* J Am Soc Mass Spectrom, 2006. **17**(11): p. 1481-9.
176. Wang, L., H. Pan, and D.L. Smith, *Hydrogen Exchange-Mass Spectrometry.* Molecular & Cellular Proteomics, 2002. **1**(2): p. 132-138.
177. Cravello, L., D. Lascoux, and E. Forest, *Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins.* Rapid Commun Mass Spectrom, 2003. **17**(21): p. 2387-93.
178. Rey, M., et al., *Recombinant immobilized rhizopuspepsin as a new tool for protein digestion in hydrogen/deuterium exchange mass spectrometry.* Rapid Commun Mass Spectrom, 2009. **23**(21): p. 3431-8.
179. Kadek, A., et al., *Aspartic protease nepenthesin-1 as a tool for digestion in hydrogen/deuterium exchange mass spectrometry.* Anal Chem, 2014. **86**(9): p. 4287-94.
180. Yang, M.L., et al., *Recombinant Nepenthesin II for Hydrogen/Deuterium Exchange Mass Spectrometry.* Analytical Chemistry, 2015. **87**(13): p. 6681-6687.
181. Bai, Y., et al., *Primary structure effects on peptide group hydrogen exchange.* Proteins, 1993. **17**(1): p. 75-86.
182. Konermann, L., A.D. Rodriguez, and M.A. Sowole, *Type 1 and Type 2 scenarios in hydrogen exchange mass spectrometry studies on protein-ligand complexes.* Analyst, 2014. **139**(23): p. 6078-6087.
183. Brown, K.A. and D.J. Wilson, *Bottom-up hydrogen deuterium exchange mass spectrometry: data analysis and interpretation.* Analyst, 2017. **142**(16): p. 2874-2886.
184. Engen, J.R., *Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS.* Anal Chem, 2009. **81**(19): p. 7870-5.
185. Chalmers, M.J., et al., *Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions.* Expert Rev Proteomics, 2011. **8**(1): p. 43-59.
186. Chen, F., et al., *High-mass matrix-assisted laser desorption ionization-mass spectrometry of integral membrane proteins and their complexes.* Anal Chem, 2013. **85**(7): p. 3483-8.
187. Chen, F., et al., *Applying mass spectrometry to study non-covalent biomolecule complexes.* Mass Spectrom Rev, 2016. **35**(1): p. 48-70.
188. Köhler, M., et al., *Binding Specificities of Nanobody•Membrane Protein Complexes Obtained from Chemical Cross-Linking and High-Mass MALDI Mass Spectrometry.* Analytical Chemistry, 2018. **90**(8): p. 5306-5313.
189. Hillenkamp, F. and J. Peter-Katalinic, *MALDI MS: a practical guide to instrumentation, methods and applications*. 2013: John Wiley & Sons.
190. Larson, R.G., *Re-Shaping the Coffee Ring.* Angewandte Chemie International Edition, 2012. **51**(11): p. 2546-2548.
191. Deegan, R.D., et al., *Capillary flow as the cause of ring stains from dried liquid drops.* Nature, 1997. **389**(6653): p. 827.
192. Wortmann, A., et al., *Investigation of the first shot phenomenon in MALDI mass spectrometry of protein complexes.* Analyst, 2007. **132**(3): p. 199-207.
193. Horneffer, V., K. Strupat, and F. Hillenkamp, *Localization of noncovalent complexes in MALDI-preparations by CLSM.* Journal of the American Society for Mass Spectrometry, 2006. **17**(11): p. 1599-1604.
194. Wenzel, R., et al., *Detector device for high mass ion detection, a method for analyzing ions of high mass and a device for selection between ion detectors*. 2013, Google Patents.

195. Farmer, T.B. and R.M. Caprioli, *Determination of protein–protein interactions by matrix-assisted laser desorption/ionization mass spectrometry.* Journal of mass spectrometry, 1998. **33**(8): p. 697-704.
196. Glocker, M.O., et al., *Characterization of specific noncovalent protein complexes by UV matrix-assisted laser desorption ionization mass spectrometry.* Journal of mass spectrometry, 1996. **31**(11): p. 1221-1227.
197. Sinz, A., *The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks.* Expert review of proteomics, 2014. **11**(6): p. 733-743.
198. Mädler, S., et al., *Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities.* Journal of Mass Spectrometry, 2009. **44**(5): p. 694-706.
199. Chen, F., S. Nielsen, and R. Zenobi, *Understanding chemical reactivity for homo- and heterobifunctional protein cross-linking agents.* J Mass Spectrom, 2013. **48**(7): p. 807-12.
200. Migneault, I., et al., *Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking.* Biotechniques, 2004. **37**(5): p. 790-802.
201. Strambio-De-Castillia, C., *SDS-PAGE sample buffer formulation.* https://openwetware.org/wiki/SDS-PAGE_sample_buffer_(Morris_formulation), 1992-2005. **original source currently not available**.
202. Liu, Y., et al., *Use of a fluorescence plate reader for measuring kinetic parameters with inner filter effect correction.* Analytical biochemistry, 1999. **267**(2): p. 331-335.
203. Morgner, N. and C.V. Robinson, *Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies.* Anal Chem, 2012. **84**(6): p. 2939-48.
204. Strohalm, M., et al., *mMass data miner: an open source alternative for mass spectrometric data analysis.* Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry, 2008. **22**(6): p. 905-908.
205. Tyanova, S., T. Temu, and J. Cox, *The MaxQuant computational platform for mass spectrometry-based shotgun proteomics.* Nat Protoc, 2016. **11**(12): p. 2301-2319.
206. Kavan, D. and P. Man, *MSTools—Web based application for visualization and presentation of HXMS data.* International Journal of Mass Spectrometry, 2011. **302**(1): p. 53-58.
207. Houde, D., S.A. Berkowitz, and J.R. Engen, *The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies.* Journal of pharmaceutical sciences, 2011. **100**(6): p. 2071-2086.
208. Gonzalez, L.S., B. Anson, and A. Mesecar, *Effect of Allosteric Changes in MERS 3CL protease Enzymatic Activity and Dimerization.* The FASEB Journal, 2017. **31**(1_supplement): p. 601.7-601.7.
209. Lorenzen, K., et al., *Determination of stoichiometry and conformational changes in the first step of the P22 tail assembly.* Journal of molecular biology, 2008. **379**(2): p. 385-396.
210. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* Journal of molecular biology, 2001. **305**(3): p. 567-580.
211. Ishida, T. and K. Kinoshita, *PrDOS: prediction of disordered protein regions from amino acid sequence.* Nucleic acids research, 2007. **35**(suppl_2): p. W460-W464.
212. Root, K., et al., *Insight into signal response of protein ions in native ESI-MS from the analysis of model mixtures of covalently linked protein oligomers.* Journal of The American Society for Mass Spectrometry, 2017. **28**(9): p. 1863-1875.
213. Dunne, M., et al., *Crystal Structure of the CTP1L Endolysin Reveals How Its Activity Is Regulated by a Secondary Translation Product.* J Biol Chem, 2016. **291**(10): p. 4882-93.
214. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Molecular systems biology, 2011. **7**(1): p. 539.

HPI Heinrich Pette Institute
Leibniz Institute for Experimental Virology

HPI • Martinistraße 52 • D-20251 Hamburg

Leibniz Association

Timothy Soh, Ph.D.
Research Department of
"Structural Cell Biology of Viruses"
Hamburg, May 26, 2019

English language certificate

Boris Krichel's Ph.D. thesis entitled **"In vitro processing and complex formation of coronavirus polyprotein NSP7-10 region"** is written in fluent English. I confirm that the language is understandable and properly articulated.

Kind regards,

Declaration of Authorship

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids. The submitted written document corresponds to the file on the storage medium. I further declare that this thesis has not been presented previously to another examination board.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich versichere, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde.